YALE ECONOMIC ESSAYS

VOLUME I NUMBER 2

Warrant Prices as Indicators of Expectations and Preferences — *Case M. Sprenkle*

Price Movements and Specie Debasement in France, 1295–1395 — *Harry A. Miskimin*

The Impact of Tightening Credit on Municipal Capital Expenditures in the United States — *Charlotte DeMonte Phelps*

A Model of Bank Portfolio Selection — *Richard C. Porter*

Price–Cost Structure and Behavior of Profit Margins — *Kazuo Sato*

FALL 1961

YALE ECONOMIC ESSAYS

This journal is intended to serve two principal functions. All doctoral dissertations in economics at Yale University will be published in its pages either in full or in extended summary. It is also a medium for economic comment and analysis by faculty, students, and graduates in economics of Yale University.

Editorial Board: Neil W. Chamberlain
Raymond P. Powell Arthur M. Okun

Volume I Fall 1961 Number 2

Yale Economic Essays is printed at the Carl Purington Rollins Printing-Office of the Yale University Press. A one-year subscription (two issues) is $4.00. Single copies are $2.50. The rate for students for one year is $2.00. Checks should be made payable to Yale University. Please address any correspondence to the Editorial Board, 37 Hillhouse Avenue, New Haven, Connecticut.

Copyright 1961 by the Yale Economic Essays.

YALE ECONOMIC ESSAYS

TABLE OF CONTENTS

Warrant Prices as Indicators of Expectations and Preferences — *Case M. Sprenkle* — 179

Price Movements and Specie Debasement in France, 1295–1395 — *Harry A. Miskimin* — 233

The Impact of Tightening Credit on Municipal Capital Expenditures in the United States — *Charlotte DeMonte Phelps* — 275

A Model of Bank Portfolio Selection — *Richard C. Porter* — 323

Price–Cost Structure and Behavior of Profit Margins — *Kazuo Sato* — 361

Summary: The purpose of this essay is to obtain empirical estimates of investors' expectations and preferences; in particular estimates of the expected price of common stocks, of the expected risk on common stock, and of the preferences of investors for taking or not taking on risk. A generalized theory of warrant prices is first developed, then a particular formula for the prices of warrants is derived on the assumption that investors think stock prices are lognormally distributed. This assumption is at least partly justified by empirical tests which show that actual stock prices are not only lognormally distributed but also follow a simple random walk. A statistical technique for estimating investors' expectations and preferences and also, incidentally, for explaining warrant prices is derived in an appendix. The data used were from the periods of the late 1920's, early 1930's and middle 1950's. The empirical results show first that the theory explains warrant prices with hardly any error. There were some difficulties in estimating expected stock prices, but the results indicate that it is usually quite close to the current price of the stock. The estimates of expected risk show that it is highly influenced by the risk of the stock in the very recent past. The empirical results also showed that even investors in the quite speculative warrant market usually have to be paid for taking on risk. Finally the results indicate that the time horizons of investors in the warrant market are not too long, and the expected returns on investing in warrants appear to be very low.

This essay was submitted to Yale University in 1960 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics at the University of Illinois.

Dissertation Committee: Professors Arthur M. Okun (Chairman), Edward C. Budd, Harold W. Watts.

*

This dissertation was written under the supervision of Arthur M. Okun and Harold W. Watts, and I would like to express my gratitude for their help and encouragement which went far beyond the call of duty. Several discussions with Arthur Okun and James Tobin led to the basic ideas contained in this study, and much of the final result can be traced to those discussions. I would like to thank John W. Hooper and T. N. Srinivasan for help in checking through some of the mathematics. The large amounts of computing time were financed by the Yale University Computing Center and Cowles Foundation for Research in Economics at Yale University. I would also like to thank my wife, Elaine, for help in matters of presentation and for typing the initial drafts. Finally, thanks are due Josephine Boron for typing the final draft.

C. M. S.

WARRANT PRICES AS INDICATORS OF EXPECTATIONS AND PREFERENCES

Case M. Sprenkle

1. Introduction

The world of uncertainty has increasingly enchanted the economist in recent years. As opposed to his old habit of relegating this world to vague footnotes and appendices, his new habit has been not only to bring it into the larger type of his discussions but to incorporate it directly into his theories and even to build new theories around the explicit recognition of this world. In the history of the development of various fields of study it is often if not usually true that the growth of new theories has been followed, with substantial lags, by their application to empirical questions. The field of uncertainty in economics has followed this rule. Although theory based on uncertainty cannot be said to be fully developed, it still is quite far advanced in relation to any empirical work done. In fact, empirical work is still almost nonexistent. The primary purpose of this study is to present some empirical work by obtaining estimates of some variables which are relevant for any theory of uncertainty.

Uncertainty affects decisions. Economists have increasingly granted explicit recognition to the fact that, since it is impossible to know for certain what will happen in the future, decisions can only be based on what is expected to happen. Thus the importance of investigating how expectations are formed no longer needs to be emphasized. But in actual fact there has been little, if any, empirical work done with expectations. What expectations do people have, how do they form them, how do past events influence them--these are questions which the economist cannot at present answer. This study attempts to obtain measures of expectations and their relations to past events.

Decisions are made by choosing among alternative courses of action which have uncertain outcomes. But these alternative courses of action have differing amounts of uncertainty or risk. For instance, an investor might have to decide between investing in a "blue chip" stock or in a highly speculative uranium mining stock. The blue chip has relatively little risk--he thinks it highly improbable that its price will change drastically, and he is quite sure that it will pay its normal dividend. The investor does not think the return on the blue chip will be very uncertain. On the other hand, he thinks the return on the mining stock is very uncertain. There is a relatively great chance of incurring very large losses. His choice will depend upon not only what he believes the average return on each to be, but also on the effect of risk considerations. If he does not like to take on risk he will be less likely to choose the uranium stock, and if he likes to take on risk he will not especially care for the relative certainty of the blue chip.

Thus a person's decisions are based not only on his expectations but also on his attitudes toward risk. A person who does not like risk, a "risk averter," will decide, among choices having the same returns, on the one with the least risk, and a person who likes risk, a "risk lover," will decide on the choice with the most risk. In between the two extremes are people who will decide on choices which have intermediate amounts of risk. If the amount of risk is of no importance to the investor in either way--he neither likes nor dislikes it--he is said to be risk neutral. At present we know very little about people's preferences toward risk. This study also attempts to obtain measures of these preferences.

Although quantitative studies of expectations and preferences are needed in all the various branches of economics, it is clearly impossible for one study to attempt to fill these needs. This study is limited to the field of asset choice, and not only is it limited to this field but it is further limited

to a study of expectations and preferences toward one particular type of asset--certain common stocks. The focus of this study is thus extremely narrow and it is made even narrower by the fact that the techniques derived here are in general not applicable to other fields nor even to a study of expectations and preferences toward other types of assets, nor even to all common stocks. Yet it should not be inferred from this that the focus of attention of this study is so narrow as to render the study utterly useless. The stock market is an extremely important and interesting market, and a study of the expectations and preferences of investors in this market would be worthwhile in its own right. But further, there is reason to believe that expectations and preferences of investors in common stocks may be indicative of those of investors in other markets or even of other kinds of expectations and preferences. There is no reason to believe that the relation between past experience and expectations in the stock market is so very different from this relation in other markets.

2. Expectations

It is necessary to understand just what expectations consist of before it is possible to develop relatively simple and efficient ways of describing and measuring them. Expectations are inherently subjective; they do not necessarily have any objective basis. Two people having exactly the same intelligence and information as to the variables affecting a certain future event will very likely have differing expectations as to the outcome. Since the world is uncertain, people's expectations are not usually formed by deciding which of the many possible outcomes will actually occur. Instead expectations will most likely consist of estimates of the likelihoods of each of the possible outcomes occurring. The person in effect assigns a probability to each of the various possible outcomes. If these were objective probabilities which gave, for instance, the chance of obtaining any number of heads in a large number of throws of a coin, we could determine the chances of obtaining any number by means of a frequency function. In the same way, we can describe expectations by means of a subjective frequency function or subjective probability distribution of expected outcomes. These distributions of expected outcomes can be of almost any size and shape or form; for instance, the outcomes can be binomially, rectangularly, normally, etc., distributed. The problem of describing and measuring these distributions consists of finding the correct form or shape of the distribution and then obtaining measures of its relevant parameters. In this study we are particularly interested in investors' subjective probability distributions of expected changes in the prices of common stocks. Reasons are given below for assuming that the form of these distributions is lognormal.

We are interested in obtaining measures of two aspects of these distributions; we need some kind of measure of central tendency and also some kind of measure of risk. For a measure of central tendency we will use the mean of the distribution and for a measure of risk we will use the variance or standard deviation.[1] Any investor would be willing to pay more

1. The explicit use of these measures and some of the implications of their use were first developed by Harry Markowitz in "Portfolio Selection," Journal of Finance, vol. VII, no. 1 (March 1952), pp. 77-91. This method of analysis has proved especially fruitful and much recent work has been based upon it. This study is, of course, heavily dependent upon this work, and grateful acknowledgement of this author's debt is hereby made. For a discussion of the advantages and disadvantages of some alternative measures of central tendency and risk, see Harry Markowitz, Portfolio Selection, Ef-

for a stock the higher the mean expected future price is. Given any mean expected future price, the higher the variance of expected prices the less would a risk averter pay for the stock but the more would a risk lover pay for it. Thus it is assumed that an investor's expectations as to future prices of common stocks can be fully described by two parameters, the mean expected price and the variance or standard deviation of expected prices. His decisions are based, then, on these two parameters and upon his preferences toward risk. The explicit purpose of this essay, then, is to obtain quantitative measures of particular investors' mean expected change in stock prices, the variance or standard deviation of changes in stock prices, and his preferences toward risk.

Since all expectations and preferences are inherently subjective, serious problems arise in trying to obtain quantitative measures of them. There are three general methods of obtaining these measures. The first is simply to ask appropriate people what they expect and what their preferences are. This is the most direct method, and it has been used with varying degrees of success in some areas, but it is usually hampered by the inability and/or unwillingness of people to quantify their estimates of their expectations and preferences. The second method is to assume that expectations are based in a particular way on what has happened in the past; for instance, expected risk might be assumed to be some function of historical risk defined in a particular way. On the basis of these assumed expectations, one could infer implicit values of preferences from decisions people actually make. This method is open to the objection that there is no way to check these assumptions so the values of the parameters obtained may be completely worthless. It has to be used in conjunction with some other method. The third method is to deduce people's expectations and preferences from their actual decisions on the assumption that they act rationally in accordance with their beliefs and preferences. The trouble with this latter method is that it is usually impossible to separate the effects of their expectations from their preferences. It is not surprising, then, that little quantitative work has been done in view of the difficulties of all these methods. However, it is possible, in theory at least, to use the latter method and separate out the effects of expected mean values, risk, and preferences at least one type of asset--certain common stocks. Basically this should be possible because there are two inter-connected assets, one of which is an option to buy or sell a particular common stock at a certain fixed price (called the exercising price) during a fixed time period. The mathematical expected value of the option depends not only on the expected mean price of the stock but also on the size and shape of that part of the distribution of expected stock prices which lies above the exercising price. Thus the price which even a risk neutral investor is willing to pay for the option depends not only on the expected mean price of the stock but also on its standard deviation. This can be most easily seen by means of the simple examples given in Section 5.

As will be explained below, by observing the actual prices paid for an option when the current price of the stock is changing, it should be possible to estimate what expectations (expected mean price and expected variance) the "market" or marginal investor must have had if it or he was investing rationally. These expectations are not those of the stock "market"

---

ficient Diversification of Investments, Cowles Foundation for Research in Economics at Yale University, (New York, John Wiley & Sons, 1959), chs. 3 and 4.

or the marginal investor in the stock market, but rather the expectations of those in the option market, which can be different.

An investor will pay a price for the option equal to its mathematical expected value as determined by his expected mean price of the stock and variance of expected stock prices only if he is risk neutral. If he is not risk neutral the price he is willing to pay for the option depends not only on its expected value but also on his preferences toward risk. In general, options have more risk than the common stock itself since a given percentage change in the price of the stock will result in a larger percentage change in the price of the option. This attribute of an option we can call leverage since it is similar to the increased risk which the stock of a highly leveraged company has in comparison with what it would have if the company were not highly leveraged. Since the concept of leverage is quite hazy, the following section will develop measures of leverage. It will be shown there that the price an investor is willing to pay for leverage can be taken as a measure of his preferences toward risk. It will be shown below that in theory it should be possible to estimate the price that investors in the option market are willing to pay for leverage, and a measure of their preferences toward risk can be obtained. Thus a study of option prices should in theory enable us to obtain quantitative measures of not only the mean price of different stocks and their expected variances expected by investors in the option market, but also quantitative measures of their preferences toward risk. Unfortunately, it was found that in general the data would not support the estimation of all three of the parameters--expected mean prices, expected variances, and preferences toward risk. In fact it was found possible to estimate only whether investors in the option market preferred to take on risk or preferred not to take on risk and the variances of the prices of the stocks expected by these investors. These estimates were made on the basis of a particular assumption as to expected mean prices. However, there are reasons for believing that the particular assumption made is not unrealistic. To offset the loss of any estimates of expected mean prices it will be shown that it is possible to obtain estimates of the rate of return or discount rate required by investors in the option market. These rates of return are particularly useful as an added indication of these investors' preferences toward risk.

3. Leverage

The concept of leverage is particularly hazy and ill defined in the literature. It is usually explained by the use of simple examples and is hardly ever defined except in terms of these particular examples. The concept of leverage always refers to some kind of asset; the asset has or contains leverage. This is true even though we speak of a "highly leveraged (or levered) company." What we really mean is that the stock of that company is highly leveraged. When we speak of leverage we must be referring to particular assets.

The most essential feature of leverage is that it increases the amount of risk or uncertainty of the asset; that is, the higher the amount of leverage an asset has, the greater the risk or uncertainty in the price of that asset.[1]

---

1. For instance, Graham and Dodd define leverage as "a quantity or attribute of a common stock produced by the presence of a substantial proportion of senior capital, and producing in turn large changes in earning power and value of the common stock in response to relatively small changes in the earning power and value of the corporation as a whole." Benjamin Graham and David L. Dodd, Security Analysis, 3rd ed., (McGraw Hill, New York, 1951), p. 556.

The asset usually considered in the simple examples is common stock. The amount of leverage per dollar of stock is usually either measured by the debt-equity ratio or the total assets-equity ratio of the company. Let us see what is implied by these measures of leverage.

We define the debt of a corporation to be the face value of all long term debt outstanding (D), equity to be the book value of all the common stock (C), and total assets to be the sum of the debt plus equity defined in this way (A). Then $A = C + D$. The debt equity ratio is D/C and its total assets equity ratio is A/C. For simplicity we assume there is no preferred stock. Suppose there is a corporation with total assets of one million dollars and a mean return of 10% per year on the million dollars. Suppose that dollar returns are lognormally distributed implying that there is an equal chance of future returns being either 11% or 9% per year, 12% or 8% per year, or in general $10 + x\%$ or $10 - x\%$ per year. Suppose further that this corporation has no debt--only equity; then the mean return on equity will also be 10% and future returns will be lognormally distributed. Let r_a = percentage return on total assets, r_s = percentage return on equity. If the current price of the stock (Ps) is proportional to r_s at all times, that is, $Ps = kr_s$ where k = price-earnings ratio and is assumed constant, then it will also be distributed lognormally if r_s is so distributed. The standard deviation of Ps will be k times the standard deviation (σ_{r_s}) of r_s. If r_s rises from 10% to 11%, Ps will rise from k (.10) to k (.11). This is the case of a nonleveraged company. Its debt-equity ratio is 0 and its total assets-equity ratio is 1. Since a company with no debt is said to be nonleveraged, it is better to use a measure of the amount of leverage which is zero in this case. The debt-equity ratio is better than the total assets-equity ratio. The amount of leverage (L) is usually considered to be $L = D/C$ or $A/C - 1$.

Now consider another corporation which has the same total assets, mean return on total assets and standard deviation of mean return. The only difference between the two corporations is that 50% of the total assets of the latter is represented by bonds which bear interest at, say, 5% a year. Then the mean return on equities is 15% instead of 10% as in the previous example--interest payments would be $25,000 per year and the mean dollar return on equity $75,000. Since the book value of equity is now $500,000, mean percentage return on equity r_s is 75/500 or 15%. Suppose that r_a rose to 11%; the dollar return on total assets rose to $110,000. Then r_s would rise to 85/500 or 17%. The same rise in r_a in the first example led to a rise in r_s of only 1% instead of 2%; thus the standard deviation of r_s is twice as great with 50% of total assets represented by equity compared with the case when 100% of net worth was represented by equity. This company is said to be leveraged. Its debt-equity ratio is 1 and its total assets-equity ratio is 2. If k is the same in both these cases then the standard deviation of Ps when $A/C = 2$ is twice as great as when $A/C = 1$. The existence of leverage leads to a higher risk on the stock, and in fact the risk on the stock is directly proportional to the amount of leverage measured by the asset-equity ratio.

This result is independent of the interest rate paid by the firm on its debt. Suppose the firm pays 10% per year instead of 5% as in the above example. Its yearly interest payments are then $50,000 and its mean dollar return on equity is $50,000 so r_s is now 10% instead of 15%. If r_a now rises from 10% to 11%, r_s will be 60/500 or 12%. Again its standard deviation will be twice as great as when $A/C = 1$. The interest rate on debt does not influence the amount of leverage nor the standard deviation of r_s.

In these simple examples the amount of leverage is best given by the

debt-equity ratio, D/C. This measure gives the amount of leverage per dollar invested; thus the total amount of leverage which a share of the stock has is given by $Ps(D/C)$. But the debt-equity ratio is only of interest because it affects returns on various assets. It does not affect r_a, but it does affect r_s. It also affects the interest charges on its debt since a large amount of debt means a greater chance of default. This factor does not affect the amount of leverage itself, but, of course, is of interest to prospective lenders. r_s is affected since its standard deviation is higher. The rise in σ_{r_s} is really the crucial factor--this is what is of interest to the investor in the corporation's stock. The amount of leverage per dollar as measured by the debt-equity ratio is really a simple proxy for the fact that σ_{r_s} is raised if there is leverage. If σ_{r_a} = the standard deviation of percentage returns on total assets, we could define the amount of leverage per dollar of stock to be $L = \frac{\sigma_{r_s}}{\sigma_{r_a}} - 1$.[1] That this measure of leverage is equivalent to the debt equity ratio is easily seen. Let R_a = dollar returns on total assets. Then $r_a = \frac{R_a}{A}$, and thus $A\sigma_{r_a} = \sigma_{R_a}$. But $r_s = \frac{R_a - D_i}{C}$ where i is the interest rate on the company's indebtedness, and thus $C\sigma_{r_s} = \sigma_{R_a}$. Therefore $\frac{A}{C} = \frac{\sigma_{r_s}}{\sigma_{r_a}}$ and the two measures of leverage are identical.

With this new measure of leverage, the concept need not be restricted to these simple examples. It can be used to compare any two assets whose returns and variance of returns are related. For instance, a person who makes a down payment on a house and takes out a mortgage takes on leverage. The returns and variance of returns on the down payment are related to those of the total value of the house. Similarly, if we want to find the leverage that an option to buy a particular stock has in comparison with the stock itself, the general formula holds. If σ_{r_1} = the standard deviation of percentage changes in the price of the option and σ_{r_2} = the standard deviation of percentage changes in the price of the stock, then the leverage which one dollar invested in the option has above and beyond any leverage the stock has is, $L = \frac{\sigma_{r_1}}{\sigma_{r_2}} - 1$. The total quantity of leverage the option has is, $Q = P\left[\frac{\sigma_{r_1}}{\sigma_{r_2}} - 1\right]$ where P = the

---

1. Modigliani and Miller seem to suggest a definition similar to this. At one point they define leverage as "financial risk." This term, however, is left vague and in the rest of their work they use the debt-equity ratio as the measure of leverage. This is legitimate for their use since they compare the leverage of stocks of corporations which they assume to have nearly equal standard deviations of percentage returns on net worth. See Franco Modigliani and M. H. Miller, "The Cost of Capital, Corporate Finance, and the Theory of Investment," American Economic Review, vol. XLVIII, no. 3, June 1958, pp. 261-297. See especially pp. 267-276.

price of the option.

If one asset has leverage in comparison to another related asset, it has more risk than the other asset. But as we have seen from the previous section, investors base their decisions on both the risk and return of an asset. If an investor likes risk, he will pay to obtain more of it. If two related assets have the same expected return but one has more risk than the other, a risk lover would pay more for the asset which has more risk. But to say that the one asset has more risk than the other is equivalent to saying that the one asset has leverage in comparison with the other. Then to say that a risk lover is willing to pay to obtain more risk is equivalent to saying that the investor is willing to pay a positive price for leverage. A risk averter, on the other hand, would only be willing to pay a negative price for leverage. We can define the price of leverage (P_e) to be the price per dollar of the asset an investor would be willing to pay for one unit of leverage. If P' is the price he would pay for the asset if he were risk neutral, then the total quantity of leverage (Q) it has is $Q = P'L$. The total amount the investor is willing to pay for the leverage this asset has (A) is $A = QP_e = P'PL_e$. He would be willing to pay this much more for the asset with the leverage than he would for the other asset. The price (P) he would be willing to pay for the asset would be $P = P' + A = P'LP_e + P' = (1 + LP_e)P'$. If the investor were a risk averter P_e would be negative, and he would be willing to pay less for the asset with leverage than he would for the other asset. In the case of stock options, P_e = the price per dollar of option the investor is willing to pay for one unit of leverage, and the total amount he is willing to pay for the leverage the option has is, $A = P'P_e\left[\frac{\sigma_{r_1}}{\sigma_{r_2}} - 1\right]$. He would then pay for the option $P = P' + A = P'\left[1 + P_e\left(\frac{\sigma_{r_1}}{\sigma_{r_2}} - 1\right)\right]$. If P_e is positive, the investor will be willing to pay more for the option than a risk neutral investor, and if P_e is negative he will be willing to pay less.

Besides his being a risk lover there is another reason an investor might be willing to pay a positive price for leverage. If it is thought that the price of the nonlevered asset will rise, then it must be thought that the price of the levered asset will rise by an even greater percentage. Thus if a risk averter thought the expected return of the leveraged asset was enough greater than the expected return of the other, even he would be willing to pay a positive price for leverage in order to gain a higher expected return. If we know that an investor thinks that the mean expected price is equal to the present price of an asset, then if he is willing to pay a positive price for leverage, he must be a risk lover and the amount he is willing to pay is a good measure of his preferences for risk. With this assumption we can infer his preferences for risk by seeing how much he is willing to pay for leverage.

Is there any limit to the amount which an investor would pay for an option? A risk averter would not pay so much for leverage that his possible gain is completely offset by his payment for leverage, but for a risk lover, of course, this sets no limit. But either type of investor may be able to obtain leverage by borrowing (buying on margin, mortgaging, etc.). (It is worth emphasizing that any investor who is willing to borrow would also be willing to pay a positive price for leverage.) The cost of leverage obtained in this manner is the interest rate at which he can borrow, and the investor would not pay more for obtaining leverage than the cost of borrowing. However, the capital market may be imperfect in the sense that the investor is unable to borrow. If this is the case there might be

no limit to the amount a risk lover would pay for leverage, and for a risk averter the only limit would be at the point where his possible gains are offset.

4. Stock Options

There are five types of stock options: rights, calls, convertible bonds, options, and warrants. They all have in common given exercising prices and given lengths of time during which they can be converted into stock by paying the exercising price. Rights are issued by the company itself, generally to existing stockholders. They are usually only exercisable for a month or two at most and are marketable. Calls are issued by private individuals and sold through specialized dealers. Most are exercisable for only six months or less and are marketable.[1] Convertible bonds are issued by the company itself. They consist of two parts which cannot be physically separated, the ordinary bond part and the option part. It is impossible to buy or sell just the option part; thus there is no directly observable price for the option. The option part is usually exercisable over reasonably long periods of time--up to ten or twenty years, and the exercising price usually varies over this period. Options are usually issued to officers and key personnel of a company as a form of payment for services. When the owner converts his options, any gain made is taxed as a capital gain, and thus his net gain is greater than if he had been paid a similar amount in the form of a salary. These options are always nonmarketable. Warrants are issued by the company itself. They are usually exercisable for reasonably long periods of time, a few of them having infinite exercising periods, and they are always marketable.

Although this study is primarily concerned with warrants, much of the theory is directly applicable to any kind of stock option. However, the empirical methods developed here depend upon the option's being marketable and having a relatively long exercising period. This means that these methods cannot be used with rights, calls, and options. They could be used with convertible bonds by separating the price paid for the bond part from that for the option part. This could be done by equating the price paid for the bond part to the price paid for bonds which are similar in every way except for being nonconvertible. Since there are very many convertible bonds at present, expectations of the future prices of a wide range of stocks can be obtained. For two reasons, however, convertible bonds are not used in this study. First, the separation of the bond part from the option part is a complicating factor. It was thought best to start with warrants which do not have this complication. Second, the computations involved in studying just a few warrants are quite large. Any additional empirical work would require very large additional amounts of computing time.

Warrants first came into large scale use in the late Twenties. Most warrants were and are issued in connection with the sale of senior issues--either bonds or preferred stock, although several have been issued in connection with common stock. By issuing warrants along with a senior issue, a company can obtain a higher price for the senior issue and/or pay lower interest charges. The issuing of warrants involves little immediate cost to the company since the exercising price is usually

1. For an interesting description of the call market and a theoretical analysis of their prices see Richard Kruizenga, Put and Call Options: a Theoretical and Market Analysis, unpublished doctoral dissertation from the Massachusetts Institute of Technology, 1956.

above the current selling price of the stock when the warrants were issued. If the price of the stock rises in the future above the exercising price, there will be a cost to the company since, in effect, it is forced to issue new equity at below what the market would bring. However, this cost is somewhat offset by the acquisition of liquid capital, which will usually come, however at just the time when the company can least use it. Warrants were and are also issued separately from other issues--in order to compensate underwriters, promoters and management, in exchange for other securities, and to sell for cash. During the Thirties the issue of new warrants stopped and many of the existing warrants expired. Others went out of existence in bankruptcies and reorganization plans. After World War II warrants began to be issued in large numbers again, and the number of new warrants issued each year has increased right up to the present. At present there are over 150 different warrants. Some companies have as many as three different kinds of warrants. Somewhat over a third of these warrants are traded on various exchanges, primarily on the Toronto, Canadian, and American Stock Exchanges. No warrants are listed on the New York Stock Exchange. The size of issue of most warrants is relatively small; only about ten warrant issues have over a million warrants outstanding, and the volume of trading in most issues is very small. Much of the recent growth in the number of warrant issues comes from Canadian companies. These companies are usually small and quite risky, many of them being uranium mining or small oil companies. U. S. companies with warrant issues do not seem to be concentrated in one or two fields as are the Canadian. Warrants have been issued by companies in almost every major field. Yet, except for some notable exceptions, most of the U. S. companies issuing warrants are small and risky as are their Canadian brothers.

It will be instructive to give an example of the relationship between the prices of a stock and the prices of its warrant over a period of time. In Figure 1, closing monthly prices for American and Foreign Power common stock and its warrant during the period from November 1929 to June 1932 are plotted. If one tried to plot a line of best fit through these points as has been done here, this line would be curved. A straight line is inadequate to describe the relation between warrant prices and stock prices. This curve seems to go close to the origin and to be asymptotic to a straight line which starts at the exercising price on the horizontal axis and goes northeastward at about a 45° angle to this axis. The use of a linear regression model to estimate statistically this relationship is obviously invalid. The relation looks more like a hyperbola than a straight line. The theory developed below will show that this relationship is not really a hyperbola, but instead a rather complex curve involving the market's expectations and preferences. By treating the price of the stock as the independent variable and the price of the warrant as the dependent variable the parameters measuring these expectations and preferences over particular periods of time can be estimated.

Treating stock prices as independent of warrant prices is not strictly correct. It has been observed that investors value retained earnings differently from dividends. To the extent that stock prices are affected by the distribution of earnings such that the more that is retained the lower the price of the stock, then there is a conflict of interest between the stockholder and the warrant holder. The warrant holder always has an interest in retaining as much of earnings as possible since this will mean higher future prices of the stock. He wants the stock to be as much of a growth stock as is possible. He does not gain as much from dividends as does the stockholder. This conflict of interest can lead to an interdependence of the two prices if the company acts in the

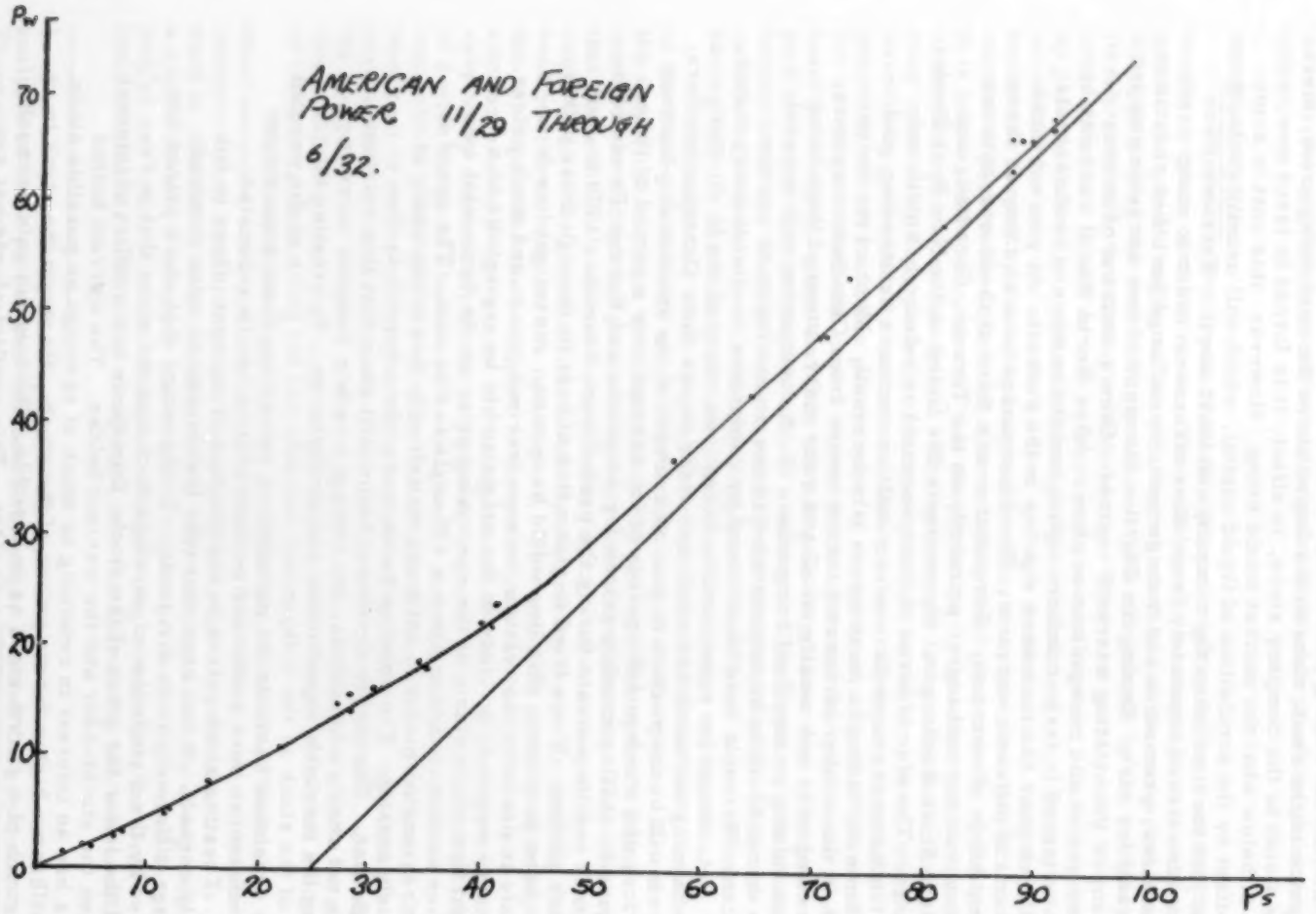

AMERICAN AND FOREIGN
POWER 11/29 THROUGH
6/32.
Pw
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
Ps

FIGURE 1.

interest of warrant holders. However, this factor would seem to be relatively minor and can be safely neglected. The most important cause of interdependence arises from exercising the warrants. This will happen only if the current price of the stock is above the exercising price since only then would warrants be exercised. However, until very close to the end of the exercising period investors will usually be able to obtain a greater return by selling their warrants on the market than by exercising them. The exception to this occurs when the price of the stock rises very high in relation to the exercising price. When this happens the price of the warrent will approach the straight line in Figure 1. This line shows the value of the warrant if it is exercised since it is a line showing the difference between the present price of the stock and the exercising price, and this is precisely what the warrant is worth when exercised. If the current market price of the warrant is close to this line an investor will not lose much by exercising his warrants. This situation has been occurring the last several years with Tri-Continental warrants which have infinite exercising periods. The price of the stock has been high in relation to the exercising price, and many warrants have been exercised each year. Even when large numbers of warrants are exercised each year, this number will usually be small in relation to the total number of shares of common outstanding and, thus, the diluting effect is apt to be small. Thus the dilution of stockholders' equity by the exercising of warrants usually will not be a threat at the end of the warrant's exercising period. If the warrants have particularly long exercising periods, this possibility of dilution will be discounted enough so that its effect on present prices should be small. By only dealing with warrants which have very long exercising periods or, better yet, indefinite (infinite) periods, this problem can be evaded.

5. Introduction to the Theory of Warrant Prices

The theory of warrant prices can best be introduced by considering several simple examples which bring out its most important and unusual characteristics. In particular they serve to show that the mathematical expected value of a warrant depends not only on the expected mean price of the stock but also on the size and shape of that part of the distribution of expected stock prices which lies above the exercising price.

To start, suppose that an investor believes that the price of a stock will remain the same indefinitely into the future. If the stock is currently selling at a price below or at the exercising price, the warrant will be worthless to him since he expects no change in the price of the stock at all. If, instead, the stock is currently selling at a price above the exercising price, he would be willing to pay the difference between the current price of the stock and the exercising price. He could buy the warrant at this price, convert it into stock at the exercising price, and pay exactly the same for the stock that he would have paid by buying the stock itself at the market price. In Figure 2 with the expected value of the warrant E[Pw] on the vertical axis and the current price of the stock (Ps) on the horizontal axis, the price he would be willing to pay for the warrant would be represented by the horizontal axis up to the conversion price (a) and the 45° line if the current price of the stock were

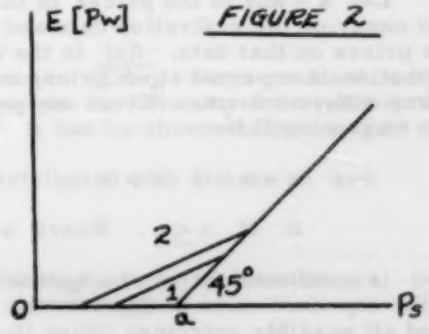


FIGURE 2

above that point.

Suppose, instead, that the stock is selling at the conversion price (a) , and the investor expected its mean future price to be the same as its present price. But he is not absolutely sure that it will remain at that price, and, in fact, he expects that there is a one-third chance it will sell at one dollar above the present price, a one-third chance it will sell at the present price, and a one-third chance it will sell at one dollar below. If the price of the stock either stays constant or falls a dollar, the warrant will be worthless, and if it rises a dollar the option will be worth a dollar. Since he believes there is a one-third chance of the stock price rising a dollar, the expected value of the option to this investor is one-third times one dollar or 33 cents. Only that part of the distribution of expected stock prices which is above the exercising price determined the warrant's expected value to the investor. If the current price of the stock is (a + 1) dollars, then it is easy to see that the possible outcomes are a + 2 , a + 1 , and a dollars, and the expected value of the warrant is one dollar. The line labeled 1 in Figure 2 shows the expected value of the warrant for various current prices of the stock. If the current stock price is outside the range (a $\pm$ 1) dollar, the expected value is the same as it was in the first example. Within this range the expected value is different in the two examples solely because the variance or standard deviation of expected stock prices is different.

Consider a third example which has the same assumptions except that the investor expects that there is a one-third chance that the price of the stock will rise by two dollars, a one-third chance it will remain constant, and a one-third chance it will fall by two dollars. The mean expected stock price is again the same as the current price. If the price of the stock actually remains constant or falls, the warrant will be worthless, but if the price of the stock rises the warrant will now be worth two dollars. The expected value of the option is now one-third times two dollars or 66 cents. The locus for this example of expected values for given current prices of the stock is shown by the line labeled two in Figure 2. It should now be apparent that all other things equal an increase in the expected risk of a stock as measured by the standard deviation or variance of expected stock prices will result in a higher expected value of the warrant.

These simple examples have ignored many complicating factors such as transactions costs, length of the suitable time period, the investor's discount rate, etc. Throughout this study trans actions costs will be neglected, and the investor's discount rate will be dealt with below. Assume, then, that the time period problem is eliminated by the fact that the warrant expires in a relatively short time, our investor planning to hold the warrant until the expiration date.

Let x = any of the prices of the stock which the investor expects might occur on the expiration date and f(x) = the distribution of expected stock prices on that date. f(x) is the investor's subjective probability distribution of expected stock prices on that date and could have any one of many different forms. Given any particular outcome of x , the value of the warrant will be

$$x - a \quad \text{if} \quad x > a$$

$$0 \quad \text{if} \quad x \leq a \,. \quad \text{Where} \quad a = \text{exercising price} \qquad (1)$$

If f(x) is continuous or can be approximated by a continuous function, then the expected value of the warrant E[Pw] to the investor will be the sum of all possible outcomes times their respectivelikelihoods of occurring

as given by $f(x)$.

$$\therefore \quad E[Pw] = \int_a^{\infty} (x - a)\, f(x)\, dx \,. \qquad (2)$$

In order to proceed much further it is necessary to make assumptions as to the form of $f(x)$. Of course, there is no way of actually knowing the form of the investor's subjective probability distribution, but there are certain a priori and empirical considerations which lead us to consider two as most promising--the normal and the lognormal distributions.

6. Random Walks

In order to describe the a priori considerations it is necessary to discuss the concept of a random walk. The kinds of random walks with which we are interested can be described as follows:

Consider a variable (x) which moves during a period of time. At the end of each small unit of time (a second, minute, or hour, etc.) during this period of time it moves and each change is independent of any previous changes. Suppose there are T units of time during the period, and the variable has an initial value of x_0 . At the end of unit one it moves and takes on a new value x_1 , and at the end of unit two it again moves to a new value x_2 , etc. The change in x is Δx . Then $x_1 - x_0 = \Delta x_1$, $x_2 - x_1 = \Delta x_2$, . . . , $x_T - x_{T-1} = \Delta x_T$. Each of the Δx_t are assumed independent of any of the other $\Delta x_1, \ldots, \Delta x_{t-1}, \Delta x_{t+1}, \ldots, \Delta x_T$ both as to absolute magnitude and to sign. Then Δx_t is a random variable which can be distributed in any way. Suppose that at the end of each unit of time the change in x is either plus one or minue one with either outcome equally likely. That is $\Delta x_t = \pm 1$. The possible outcomes (and their probabilities of cocurring in parentheses) over successive units of time can be shown as:

time (t) =	0	1	2	3	. . .
				$x_0 + 3(1/8)$	
			$x_0 + 2(1/4)$		
		$x_0 + 1(1/2)$		$x_0 + 1(3/8)$	
	x_0		x_0 (1/2)		. . .
		$x_0 - 1(1/2)$		$x_0 - 1(3/8)$	
			$x_0 - 2(1/4)$		
				$x_0 - 3(1/8)$	
$\sum_{i=1}^{t} \Delta x_t$	= 0	0	0	0	. . .
σ_t^2	= 0	1	2	3	. . .

As is shown the mean of the sum of changes $(\overline{\Sigma \Delta x_t})$ stays constant, and the variance of the sum (σ_t^2) increases as t ; thus σ_t increases as $\sqrt{t}$. If $E[\Delta x_t] = 0$ and $E[\Delta x_t]^2 = \sigma^2$,[1] it can be shown that as t gets large the sum $\sum_{t=1}^{T} \Delta x_t$ is normally distributed with a mean 0 and

1. The notation E[ ] will be used throughout to refer to the expected value of whatever variate is in the brackets.

variance $T\sigma^2$ and that x_T is also normally distributed with a mean x_0 and variance $T\sigma^2$. This can be shown as follows:

$$\text{Since } \Delta x_t = x_t - x_{t-1}$$

$$\sum_{t=1}^{T} \Delta x_t = \sum_{t=1}^{T} (x_t - x_{t-1}) \cong \int_{x_0}^{x_T} dx = x_T - x_0$$

$$\therefore x_T = x_0 + \Delta x_1 + \ldots + \Delta x_T$$

Since the Δx_t are independent, by the additive form of the central limit theorem x_T and $\sum_{t=1}^{T} \Delta x_t$ are asymptotically normally distributed. The Δx_t can be distributed in any manner consistent with the central limit theorem (e. g., they can be distributed binomially, rectangularly, normally or lognormally).[1] The Δx_t are changes over very small units of time. If changes were observed over longer periods of time, say a week or a month, we would have observations of $\sum_{t=1}^{n} \Delta x_t$, $\sum_{t=n+1}^{2n} \Delta x_t, \ldots, \sum_{t=kn+1}^{kn+n} \Delta x_t$. It has been shown that $\sum_{t=1}^{T} \Delta x_t$ is asymptotically normally distributed, thus the changes $x_{kn} - x_{kn-n}$ will also be asymptotically normally distributed.

This model corresponds exactly to the case of an investor who thinks it equally likely that the price of a stock will rise some dollar amount or fall the same dollar amount in a short period of time--say a day or an hour. An investor with this type of expectations in effect has a subjective probability distribution of expected stock prices at any time (t) which is normal. The model can be easily modified to take into account the fact that an investor might not think the mean expected price at any time is equal to the present price. He might think instead that $x_{t+1} = x_t + k + \Delta x_t$ where k is a constant and Δx_t is again an independent random movement. The expected mean price of the stock at time t is now $x_t = x_0 + kt$. In the same way that we developed a normal distribution of expected stock prices at time t around x_0 we can develop the same normal distribution of stock prices around $x_0 + kt$. The effect of k can be eliminated by taking out the trend.

We have seen that a normal distribution of expected stock prices at time t can be rationalized if the investor thinks that equal dollar rises or drops are equally likely over small units of time. It will now be shown that a lognormal distribution of expected stock prices can be rationalized if the investor thinks that the size of this rise or drop depends on the price of the stock; that is, it is percentage rises or drops which are important. In this case

$$x_t - x_{t-1} = \Delta x_t x_{t-1}$$

---

1. See William Feller, An Introduction to Probability Theory and its Applications, vol. 1, 2nd ed., (New York, John Wiley & Sons, 1957), pp. 324-5. And J. Aitchison and J. A. C. Brown, The Lognormal Distribution, Cambridge University Press, (Cambridge, 1957), p. 13 and pp. 22-24.

$$\text{or} \quad \Delta x_t \approx \frac{x_t - x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - 1 .$$

$$\text{But} \quad \sum_{t=1}^{T} \left(\frac{x_t - x_{t-1}}{x_{t-1}} \right) = \sum_{t=1}^{T} \Delta x_t \approx \int_{x_0}^{x_T} \frac{dx}{x} = \ln x_T - \ln x_0$$

if the effect at each step is small.

$$\ln x_T = \ln x_0 + \Delta x_1 + \ldots + \Delta x_T .$$

Again by the additive form of the central-limit theorem $\ln x_T$ is asymptotically normally distributed and hence, x_T and $\prod_{t=1}^{T}(1 + \Delta x_t)$ are asymptotically lognormally distributed. Since $\prod_{t=1}^{T} (1 + \Delta x_t)$ is asymptotically lognormally distributed, $\frac{x_T}{x_0}$ must also be so distributed. Thus if equal percentage rises or drops are equally likely over small units of time, then the price of the stock at any time in the reasonably far future (x_T) is lognormally distributed, the product of all the changes $\prod_{t=1}^{T} (1 + \Delta x_t)$ is lognormally distributed, and thus the ratio of the price in the future to the present price $\frac{x_T}{x_0}$ is also lognormally distributed. If we split the long period T into n sub-periods where n is not too large, the ratio of $\frac{x_n}{x_{n-1}}$ will also be asymptotically lognormally distributed. If these sub-periods are say weeks or months, the observations of these ratios over a longer period, say a year, would be asymptotically lognormally distributed.

Let $\frac{x_t}{x_{t-1}} = x'_t$. If $E[\ln x'_t] = \mu$ and $E[\ln x'_t - \mu]^2 = \sigma^2$, then the product $\prod_{t=1}^{T} x'_t$ is asymptotically lognormally distributed with a mean $T\mu$ and a variance $T\sigma^2$. A variate of this type is said to follow the law of proportionate effect.[1]

It is important to note that in this case both μ_t and σ_t^2 depend upon t . Since σ_t^2 increases as t , σ_t must increase as $\sqrt{t}$. In the previous case of a normally distributed variate it was also true that σ increases as $\sqrt{t}$, but the mean x'_t was independent of time (unless, of course, there is a trend factor k) . The expected value of x'_t when x'_t is lognormally distributed is given by

$$E[x'_t] = e^{\mu + 1/2\sigma^2} \quad \text{and} \quad E \prod_{t=1}^{T} [x'_t] = e^{t(\mu + 1/2\sigma^2)} .$$

In order to keep $E \prod_{t=1}^{T} [x'_t]$ constant over time μ_t must decrease since

1. Ibid.

σ_t^2 increases; thus, the mean of the logs of the ratios must continually decrease as t gets large. In order to keep the mean expected price equal over time (so that the stock is a "fair bet"), $E\prod_{t\equiv 1}^{T}[x_t']$ must equal one. Since $\sigma^2 > 0$, this means $\mu < 0$ but since $\ln x_t'$ is distributed normally is also equal to the mode or most likely $\ln x_t'$. But if the mode of $\ln x_t'$ is negative, the mode of x_t' must be less than one. Therefore, if a stock is distributed lognormally and is a "fair bet," the most likely outcome is that the stock will fall in price--that $\frac{x_t}{x_{t-1}} < 1$, and the chances of getting equal percentage gains or losses are not equal. When $t = \infty$, $\sigma_t^2 = \infty$, and thus $\mu_t = -\infty$ if $E\prod_{t=1}^{T}[x_t']$ is to be constant.

As a simple example of such a random walk let $x_0 = 1$ and the probability of getting either $x_t' = 1 \pm a$ equal to $1/2$. Let $1 + a = e^b$ and $1 - a = e^{-c}$ where $b \neq c$ for all a.

Then the random walk is (with the probabilities in parentheses):

time t = 0	1	2	3	. . .
			e^{3b} (1/8)	
		e^{2b} (1/4)		
	e^{b} (1/2)		e^{2b-c}(3/8)	
1		e^{b-c}(1/2)		. . .
	e^{-c}(1/2)		e^{b-2c}(3/8)	
		e^{-2c}(1/4)		
			e^{-3c} (1/8)	

Taking logs of this walk:

t = 0	1	2	3	. . .
			3b (1/8)	
		2b (1/4)		
	b (1/2)		2b-c(3/8)	
0		b-c(1/2)		. . .
	-c (1/2)		b-2c(3/8)	
		-2c(1/4)		
			-3c (1/8)	
μ_t = 0	1/2(b-c)	(b-c)	3/2 (b-c)	. . .
σ_t^2 = 0	$1/4(b+c)^2$	$1/2(b+c)^2$	$3/4(b+c)^2$	. . .

As can be seen μ_t decreases and σ_t^2 increases as t, and thus σ_t increases as $\sqrt{t}$. This model can be modified to take account of the fact that an investor might not think the mean expected price at any time is equal to the present price. He might expect instead that $x_t = (x_t')kx_{t-1}$. Then $x_t' = \frac{x_t}{kx_{t-1}}$. The expected mean price at time t, $\bar{x}_t = x_0 k^t(1+\Delta x_1)\ldots(1+\Delta x_t)$ and $\ln \bar{x}_t = \ln x_0 + kt + \Delta x_1 + \ldots + \Delta x_t$. In the same way that

we developed a lognormal distribution of expected stock prices at time t around μ_t, we can develop the same lognormal distribution of stock prices around $\mu_t k^t$. Again the effect of k can be eliminated by removing the trend.

We have now developed rationales for both normally distributed expected stock prices and lognormally distributed expected stock prices. Is there any a priori reason for preferring one to the other? The answer is yes, there is a quite important reason for preferring the lognormal distribution. Normal distributions imply that the investor thinks there is a chance that future stock prices will be negative. As t increases the variance of x_t increases until it becomes ∞, but long before this time σ_t^2 will be so large that the chances of getting $x_t < 0$ will become very large. The smaller x_0 is, the greater will the chances be, and as $x_0 \to 0$ the chances of negative x_t approach one-half. Thus for stocks with low prices relative to σ the assumption of normally distributed stock prices is especially poor. Using a lognormal distribution the variance ($t\sigma^2$) of $\ln x_t$ goes to ∞. The range of possible $\ln x_t$ is thus $-\infty$ to ∞, thus the possible range of x_t is 0 to ∞. The lognormal distribution restricts x_t to positive values, and an investor with this type of expectations does not think there is any chance of negative stock prices. A typical lognormal frequency function is shown in Figures 3a and 3b. As is seen when drawn on x_t it is skewed to the right and is bounded by 0, but when drawn on $\ln x_t$ it is normal and unskewed and is not bounded.

FIGURE 3a

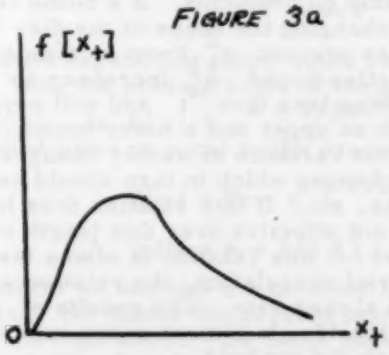
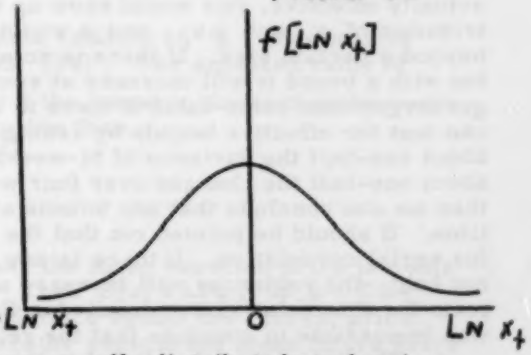


FIGURE 3b



One further implication of using normally distributed stock prices is that when the current price of the stock changes, the chances of given dollar changes stay the same if the investors estimate of the risk of the stock stays constant. For instance, when a stock is selling at, say, two dollars the chance of it rising or falling one dollar intuitively seems less than when it is selling at, say, twenty dollars. It would seem highly likely that the investor's estimate of risk would change if he thought stock prices to be normally distributed. If the investor has lognormal expectations, his estimate of percentage gains or losses would not necessarily change if the current price changes. Thus, over periods where stock prices do change substantially it seems better to assume the investor's estimate of the chances of given percentage gains or losses stays constant.

One would expect that over reasonably long periods of time investors are not fooled by the market. It has been shown above that if actual weekly or monthly changes in stock prices are normally distributed, then the distribution of possible stock prices on any future date is normal, but if the ratios of prices one month to prices the previous month are lognormally

distributed, the possible prices at any future date will be lognormally distributed. If stock prices actually follow a normal random walk, then it would seem reasonable to assume that investors think future prices to be normally distributed, but if stock prices actually follow a lognormal random walk, then the assumption of lognormal expectations would seem in order. Tests for normality and lognormality were conducted on the stocks used in this study. The results of these tests are detailed in Section 9. They show that in most cases a lognormal distribution is a better assumption than a normal distribution.

Since any form of a random walk model depends upon independence of the successive Δx_t, if they are found to be correlated, the model generally cannot be used. If it is true that the empirical Δx_t are correlated and investors are not fooled, they cannot have expectations which depend upon random walks. Ordinary serial correlation methods were used to test for independence of the Δx_t. These results are detailed in Section 10. In general they show that the amount of serial correlation is almost always statistically insignificant, and even where significant the absolute magnitude of it is small enough to be neglected.[1]

Besides certain types of non-independence of the Δx_t, another possible reason for not using simple random walks is that x_t might have upper and/or lower bounds. For instance, the investor might think that x_t cannot be greater than some finite value, say x_H. Since both the lognormal distribution and the normal distribution imply that values of x_t greater than x_H are possible, they are at best approximations in this case--and bad approximations if x_H is relatively small. If x_H is quite large in relation to the present price, then this case presents no problems. If a bound is actually effective, this would show up by changing the shape of the distribution of x_t and Δx_t, and it would also prevent $\sigma^2_{t_2}$ from increasing beyond a certain size. If there is no effective bound $\sigma^2_{t_2}$ increases as t, but with a bound it will increase at something less than t and will never get larger than some value if there is both an upper and a lower bound. We can test for effective bounds by seeing if the variance of weekly changes is about one-half the variance of bi-weekly changes which in turn should be about one-half the changes over four weeks, etc. If this relation does hold, then we can conclude that any bounds are not effective over this length of time. It should be pointed out that the test for this relation is also a test for serial correlation. If there is any serial correlation, the relation will not hold--the variances will increase at a slower rate. The results of tests for this relation are detailed in Section 11. In almost every case it was impossible to conclude that the relation did not hold.

By the method developed in this essay we can get an independent estimate of an investor's estimate of the variance of changes in stock

---

1. These results agree with other recent studies of stock prices, which come to the conclusion that both stock indices and individual stock prices behave much like wandering series or random walks with very little significant serial correlation and still less which is high in absolute magnitude. See M. S. Kendall, "The Analysis of Economic Time Series--Part I: Prices," Journal of the Royal Statistical Society, Series A, vol. CXVI, part 1, (1953), pp. 11-25; Harry V. Roberts, "Stock-Market 'Patterns' and Financial Analysis: Methodological Suggestions," Journal of Finance, vol. XIV, no. 1, (March 1959), pp. 1-10; and M. F. M. Osborne, "Brownian Motion in the Stock Market," Operations Research, vol. 7, no. 2, (March-April 1959), pp. 145-173. These studies also agree with the choice of lognormal over normal distributions of changes in stock prices.

prices at time t . Suppose this variance is m times the empirical yearly variance of changes in the price of the stock. If there were no bounds, then the investor's estimate of the variance will equal the actual variance of changes over periods of m years duration. A discussion of the implications of this relation must wait until equations for expected values and prices of warrants have been derived.

7. Expected Values and Prices of Warrants

For the reasons described in the preceding section, we will assume that an investor thinks stock prices follow a simple random walk leading to a lognormal distribution of stock prices at any time t .[1] Equation (2) above gives the expected value of a warrant to an investor,

$$E[Pw] = \int_a^{\infty} (x-a)f(x)dx \, . \qquad (2)$$

f(x) is lognormal with a variance of lnx equal to σ^2 and a mean of lnx equal to μ .
Then,

$$f(x) = \frac{1}{x\,\sigma\,\sqrt{2\pi}} \; e^{-1/2\left[\frac{\ln x - \mu}{\sigma}\right]^2} \qquad 0 \leq x \leq \infty \qquad (3)$$

and

$$E[Pw] = \int_a^{\infty} \frac{(x-a)}{x\,\sigma\sqrt{2}} \; e^{-1/2\left[\frac{\ln x - \mu}{\sigma}\right]^2} dx \, . \qquad (4)$$

We would expect that there would be some relation of μ , the mean of lnx , with the present price of the stock (Ps) . The expected value of x , E[x] , is $E[x] = e^{\mu + 1/2\sigma^2}$. If the investor thinks that the mean expected price is equal to the present price then,

$$E[x] = Ps = e^{\mu + 1/2\,\sigma^2}$$

$$\therefore \ln Ps = \mu + 1/2\,\sigma^2$$

In general we assume the investor thinks the mean expected price is some constant (k) times the present price. Then $E[x] = kPs$. If $k = 1$ we have the case above--the mean expected price equals the current price. If $k = 1.05$, for example, then the investor expects the mean future price to be 5% greater than the present price of the stock. More complex relations could be set up but would only serve to complicate matters. When $k = 1$ this relation is incorporated in the assumption that the investor has unitary elasticity of expectations since a change in the present price results in an equal change in the mean expected price.[2] Any change ΔPs in the present price of the stock then results in a change of $k\Delta Ps$ in the mean expected price of the stock, but it is assumed that σ is unchanged. The investors estimate of the variance of a stock remains unchanged with

---

1. The mathematics of the derivations obtained here are contained in Appendix A. Only the end results of the derivations are contained here.

2. The normal Hicksian elasticity of expectations implies that $E[x] = a + kPs$; our relation is the special case where $a = 0$.

changes in the present price.

Using this relation then

$$E[x] = kPs = e^{\mu + 1/2\sigma^2}$$

$$\therefore \quad \mu = \ln Ps + \ln k - 1/2\sigma^2 \tag{5}$$

Given a particular present price of the stock and substituting (5) into (4), the expected value of the warrant to the investor will be

$$E[Pw;Ps] = \int_a^\infty \frac{(x-a)}{x\sigma\sqrt{2\pi}} e^{-1/2\left[\frac{\ln x - \ln Ps - \ln k + 1/2\sigma^2}{\sigma}\right]^2} dx \tag{6}$$

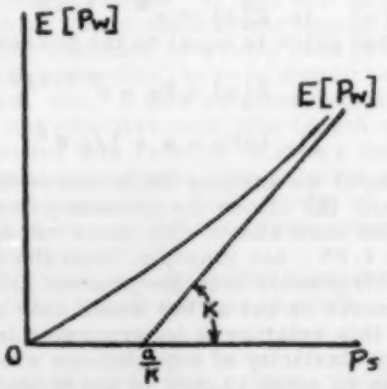
This expression can be reduced to the form,

$$E[Pw;Ps] = \frac{kPs}{\sqrt{2\pi}} \int_{-1/2\sigma - \frac{\ln Ps + \ln k - \ln a}{\sigma}}^\infty e^{-1/2\mu^2} d\mu$$

$$- \frac{a}{\sqrt{2\pi}} \int_{+1/2\sigma - \frac{\ln Ps + \ln k - \ln a}{\sigma}}^\infty e^{-1/2\mu^2} d\mu . \tag{7}$$

Both of the integrals are standardized normal cumulative frequency functions so that once the limits are known their values are easily obtained from tables. The relationship of the expected value of a warrant for different present prices of the stock is shown in Figure 4. The relation is nonlinear. When $Ps = 0$, $E[Pw] = 0$, and when Ps is very large $E[Pw]$ approaches the straight line with a slope of k and an intercept on the Ps axis of $\frac{a}{k}$. Since no matter how large Ps is, with this distribution there is always a chance that the future price of the stock will fall below a; thus the expected value of the warrant is tangent to the asymptote at $Ps = \infty$. The larger k is the higher $E[Pw]$ will be since the slope of the asymptote will be higher and its intercept smaller. The lower a is, of course, the higher $E[Pw]$ is. A higher σ^2 will also mean a higher $E[Pw]$ as it did for the simple examples in Section 5. If $\sigma^2 = \infty$, which it would for an investor with an infinite time horizon and simple random walk expectations, then $E[Pw;Ps] = kPs$. The expected value of the warrant would equal what he would pay for the stock. Since, of course, this situation never arises, we can conclude that investors either do not have infinite time horizons or they do not believe in simple random walks.

FIGURE 4



We have now developed the expected value of a warrant to an investor who has certain assumed expectations as to stock prices. But it is in general not true that the investor would be willing to pay a price for the warrant exactly equal to the expected value of it to him. In fact he would

be willing to pay exactly this price only if he were neutral to risk. If he is a "risk lover" he will be willing to pay something to take on risk. In general warrants have more risk than stock since a given percentage change in the price of the stock will result in a larger percentage change in the price of the option. This attribute of any asset was called leverage in Section 3, and the quantity of leverage per dollar of the asset price was defined to be

$L = \frac{\sigma_{r_1}}{\sigma_{r_2}} - 1$, where σ_{r_1} = standard deviation of percentage returns on one asset and σ_{r_2} = standard deviation of percentage returns on the other, and $\sigma_{r_1} > \sigma_{r_2}$.

In Section 3 we did not apply this formula to warrants and for good reasons. In the case of a highly leveraged company, returns on common stock and returns on the total assets of the company are distributed in the same way; that is, if the returns on total assets are lognormally distributed, then the returns on the common stock are also distributed lognormally. This is so because the relation between the two returns is linear; any given change in returns on total assets results in a change in returns on common stock which is a constant times the former change. Thus the standard deviation of returns on stock is also a constant times the standard deviation of returns on total assets.

The returns on warrants come through price changes; therefore, to find the leverage which warrants have, we want to use the standard deviation of percentage changes in warrant prices and the standard deviation of percentage changes in stock prices. Then the amount of leverage which one dollar of a warrant has is:

$L = \frac{\sigma_{P_w}}{\sigma_{P_s}} - 1$ where σ_{P_w} = standard deviation of prices and σ_{P_s} = standard deviation of stock prices.

However, the relation between Pw and Ps is not linear as it is in the above case, and thus Ps and Pw will be distributed in different ways; and, in fact, Pw itself will be distributed in different ways depending upon the current Ps. An explicit expression for σ_{P_w} would be very complex. Even if the distribution of Ps were symmetrical, the distribution of Pw would not be, and σ_{P_w} might not be such a good measure of risk in this case. As an alternative to the above measure, we will define the leverage per dollar of warrant to be:

$$L = \frac{\frac{dPw}{Pw}}{\frac{dPs}{Ps}} - 1 = \left(\frac{dPw}{dPs}\right)\left(\frac{Ps}{Pw}\right) - 1 .^{1} \qquad (8)$$

1. This definition of leverage for warrants was first used in Guynemer Giguere, "Warrants, A Mathematical Method of Evaluation," The Analysts Journal, vol. 14, no. 5, (November 1958), pp. 17-25.

The leverage per dollar of a warrant is thus defined to be the percentage change in the price of the warrant brought about by some percentage change in the price of the stock minus one. Or, alternatively, it is the slope of the curve showing the relation between warrant and stock prices times the ratio of the stock price to the warrant price, all minus one. If the relation between stock and warrant prices were linear, this measure would yield the same values of leverage as the measure which uses the standard deviations of returns. And for all assets which are linearly related the two measures would be identical. In the case of warrants the smaller the anticipated changes in stock prices, the closer will the two measures be. We will see that this new measure adds few complications to the theory of warrant prices. This measure is the quantity of leverage one dollar of a warrant has above and beyond whatever leverage the stock itself might have. The quantity of leverage any particular warrant has is:

$$Q = \left[\left(\frac{dPw}{dPs}\right)\left(\frac{Ps}{Pw}\right) - 1\right] \quad Pw \,. \tag{9}$$

P_e was defined as the price per unit of leverage that an investor is willing to pay for leverage. The amount an investor is willing to pay for the leverage that a warrant has is:

$$A = \left[\left(\frac{dPw}{dPs}\right)\left(\frac{Ps}{Pw}\right) - 1\right] PwP_e = \left[\left(\frac{dPw}{dPs}\right) Ps - Pw\right] P_e . \tag{10}$$

As we have seen in Section 3, there are two reasons why an investor would be willing to pay a non-zero price for leverage, either (1) he believes the mean expected price of the stock will be different from the present price of the stock, or (2) he is not neutral to risk. (Implicit in both these reasons is the assumption that the investor has limited capital and is faced by an imperfect capital market making it impossible for him to borrow.) The effects of (1) are already accounted for by the variable k which relates the mean expected price of the stock to the present price; thus a non-zero P_e can only be attributed to a non neutrality toward risk. If P_e is positive, the investor is willing to pay to take on risk--he is a "risk lover," and if P_e is negative, he has to be paid to take on additional risk--he is a "risk averter." Thus P_e in effect measures the investor's preferences toward risk. The price the investor is willing to pay for the warrant or the value of the warrant to the investor (V) is the expected value of the warrant plus the amount he is willing to pay for the leverage of the warrant, i.e.:

$$V = E[Pw;Ps] + \left[\left(\frac{dPw}{dPs}\right) Ps - Pw\right] P_e \,. \tag{11}$$

If P_e is reasonably close to zero we can approximate $\frac{dPw}{dPs}$ by $\frac{dE[Pw;Ps]}{dPs}$ and Pw by $E[Pw;Ps]$. The error introduced by this approximation will be quite small if we assume that $V = Pw$; that is, that the value of the warrant to the investor is equal to the present price of the warrant. For the investor we are most interested in, i.e., the marginal investor, this assumption holds true. For this case if $P_e > 0$, $E[Pw;Ps] \leq Pw$;

and $\frac{dV}{dPs} \geq \frac{dE[Pw;Ps]}{dPs}$ since $\frac{dPw}{dPs} \geq 0$ and $\frac{d^2Pw}{dPs^2} \geq 0$.

Thus the errors introduced by the substitutions tend to cancel each other out. They also tend to cancel each other out for the case where $P_e < 0$. However, the closer P_e is to zero the better the approximation will be. With these substitutions,

$$V = E[Pw;Ps] + \left[\frac{dE[Pw;Ps]}{dPs} \quad Ps - E[Pw;Ps]\right] P_e$$

$$= (1-P_e)\, E[Pw;Ps] + \frac{dE[Pw;Ps]}{dPs} \quad PsP_e \quad . \qquad (12)$$

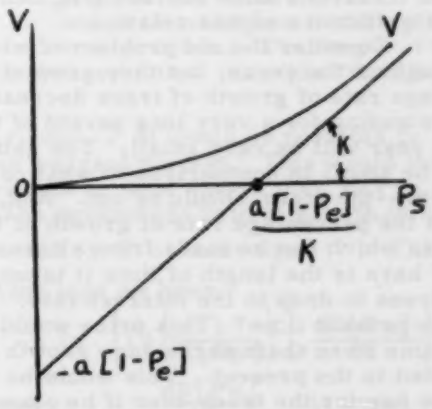
Explicit expressions for the value of a warrant to an investor or the price he is willing to pay for it can now be derived. If the investor thinks future stock prices are lognormally distributed, then:

$$V = \frac{kPs}{\sqrt{2\pi}} \int_{-1/2\sigma - \frac{\ln Ps + \ln k - \ln a}{\sigma}}^{\infty} e^{-1/2\,\mu^2} d\mu$$

$$- \frac{(1-P_e)a}{\sqrt{2\pi}} \int_{+1/2\sigma - \frac{\ln Ps + \ln k - \ln a}{\sigma}}^{\infty} e^{-1/2\,\mu^2} d\mu \, . \qquad (13)$$

The only difference between this and the expression for the expected value of the warrant is that the second term is multiplied by the factor $(1-P_e)$. As is to be expected if $P_e > 0$, $V > E[Pw]$ and if $P_e < 0$, $V < E[Pw]$. As is seen in Figure 5, V is asymptotic to a line which still has a slope of k, but the intercept of the asymptote with the axis is $\frac{a(1-P_e)}{k}$ instead of just $\frac{a}{k}$.

FIGURE 5



We have now developed expressions for the expected value of a warrant to an investor and the price he is willing to pay for the warrant given the present price of the stock. These expressions could apply to any investor. The marginal investor is just ready to either buy or sell on the market. If he buys he presumably thinks that V is somewhat greater than the actual price of the warrant (Pw), but the marginal investor who sells to him presumably thinks that V is somewhat less than Pw. The consensus of opinion between the two is that V must be close to Pw. For each individual sale the actual investors who buy and sell will usually be different from those who entered into the previous sale, but the consensus at each sale is that V must be close to Pw. Actual prices of the warrant then reflect the consensus of marginal investors' opinions--the marginal investors' expectations and preferences are the same as the market's expectations and preferences.

The primary purpose of this essay is to develop empirical measures of expectations and preferences. Actual prices paid for warrants can be treated as observations of the price the marginal investor is willing to pay, or, alternatively, the price the market will pay. Assume $Pw = V + \epsilon$ where ϵ is an error term. By taking periods of time during which the market's expectations and preferences as measured by σ, k and P_e can be assumed to remain relatively constant, we can treat this as a regression problem with Ps the independent variable and Pw the dependent

variable, and we can obtain measures of the parameters, σ, k, and P_e. Periods of time during which these parameters can be considered relatively constant are those in which there is a predominant movement in the market either up or down. During these periods the amount of optimism or pessimism in the market will be relatively constant, and this should mean that expectations and preferences will be relatively constant. During periods of time which contain strong movements in both directions, the market is apt to change its expectations or its preferences or both. For instance, during upswings one might expect that investors would be more willing to take on risk than in downswings, and P_e would then tend to be higher.[1]

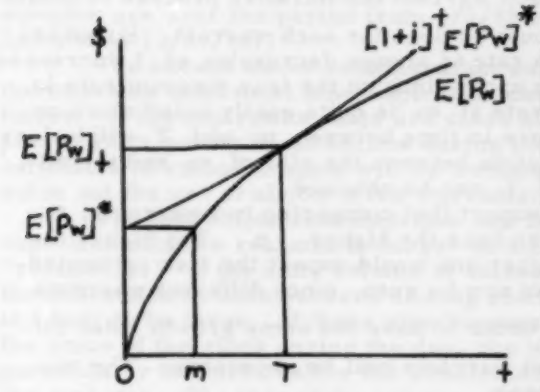
Before turning to the results it will be best to analyze some of the implications of expected values and prices of warrants. We have seen that the larger σ is the higher will the price of the warrant be, and we have also seen that if stock prices follow a random walk the observed σ's will increase as $\sqrt{t}$. Assuming a simple random walk the σ estimated through warrant prices then corresponds to the actual σ of changes in the stock price over some number (m) of years. The greater m is, the greater is the expected value of the warrant. If this m years could be considered a measure of "time horizon," then the expected value of a warrant increases as an investor's time horizon lengthens. An analogy may help to bring out the significance of this relation.

Consider the old problem of when to cut the trees. Trees grow throughout the years, but they grow at a decreasing rate; that is, the percentage rate of growth of trees decreases with age. If the trees have been left standing for a very long period of time the rate of growth over the next year will be very small. The gain from keeping the trees standing will be small in comparison to what could be earned by investing in other assets--the trees should be cut. And, of course, the trees should be cut when the percentage rate of growth of the trees is equal to the rate of return which can be made from alternative investments. The "time horizon" here is the length of time it takes the percentage rate of growth of the trees to drop to the interest rate. What would be the price of the trees at the present time? This price would simply be the value of the trees at the time when their percentage growth rate equaled the interest rate discounted to the present. This would be the price an investor would be willing to pay for the trees even if he planned to resell the trees in a very short time. His "time horizon" as used here is not the same as the length of the period he plans to hold the asset.

Turning again to warrants, we have not taken into account any discount rate. What we have obtained is the present value of the warrant $E[Pw]^*$ and a relation between the estimated σ and the σ of weekly changes in the stock prices (σ_1). The number of years given by the relation $\frac{\sigma^2}{52\,\sigma_1^2}$ is not quite the same as the "time horizon" of the tree problem. This can be seen in Figure 6. For the moment assuming that $P_e = 0$ and $k = 1$, the $E[Pw]$ in dollars of a particular warrant is shown on the vertical axis and time in years on the horizontal. $E[Pw]$ should really be interpreted as the expected exercising value of the warrant at any number of years in the future. But if over any number of

1. The statistical techniques used to estimate these parameters are described in Appendix B, and the empirical results obtained are detailed below.

FIGURE 6



years the most likely outcome occurs, that is Ps remains constant, the actual exercising value of the warrant will be the same as it was at the beginning. Over the years the trees grew in value, but most likely the warrant did not. The paradox is easily explained when it is remembered that the most likely outcome (the mode) is not necessarily the same as the average outcome, and the expected value or average outcome is what is relevant. If the rate of interest is i then $(1+i)^t\ E[Pw]^*$ shows the value at any time in the future of $E[Pw]^*$ dollars invested now. Where it is tangent to the expected exercising value of the warrant, it gives the "time horizon" T which is analogous to that of the tree problem. The length of time m implies by the ratio $\frac{\sigma^2}{52\,\sigma_1^2}$ will always be less than this "time horizon." If one could obtain the "time horizon" T it would be possible to obtain the relevant discount rate for the warrant market. This would involve an iterative process but could easily enough be done except that $E[Pw]$ for each warrant will be quite different. The percentage rate of growth of a warrant is $\left(\frac{1}{E[Pw]}\right)\left(\frac{dE[Pw]}{dT}\right)$ and at T it must equal i. Therefore $E[Pw] = E[Pw]^*\ (1+i)^t$. Taking logs we have

$$\ln E[Pw]^* = \ln E[Pw] - t \ln \left[1 + \left(\frac{1}{E[Pw]}\right)\left(\frac{dE[Pw]}{dT}\right)\right]$$

We can always put t in terms of σ and σ_1. $\frac{dE[Pw]}{dT}$ is equal to $\left[\frac{dE[Pw]}{d\sigma}\right]\left[\frac{d\sigma}{dT}\right]$ where $\frac{d\sigma}{dT} = \frac{26\sigma_1^2}{\sigma}$. Then

$$\ln E[Pw]^* = \ln E[Pw] - \frac{\sigma^2}{52\sigma_1^2} \ln \left[1 + \left(\frac{1}{E[Pw]}\right)\left(\frac{dE[Pw]}{d\sigma}\right)\frac{26\sigma_1^2}{\sigma}\right].$$

This equation will hold only at T, and solving it is the only method of obtaining T. The left side of the equation is known, and the right side is a function solely of σ since σ_1 and $\frac{Ps}{a}$ are given for each warrant. But it is impossible to solve the equation explicitly for σ so an iterative method of solution is necessary. But for a lognormal distribution of stock prices $E[Pw]$ and $\frac{dE[Pw]}{d\sigma}$ and thus the percentage rate of growth of each warrant, depend not only on σ but on $\frac{Ps}{a}$ and σ_1. At any any given t (or given σ and σ_1) the higher $\frac{Ps}{a}$ is, the lower will

be the growth rate. At any given t for $\frac{Ps}{a} < 1$ the lower σ_1 is, the higher the growth rate, but for $\frac{Ps}{a} > 1$, the lower the growth rate. Since $\frac{Ps}{a}$ and $_1$ are different for each warrant the iterative process to obtain the relevant growth rate and T must be done for each warrant. However, over the relevant range the growth rate is always decreasing as t increases, thus the growth rate at m sets an upper limit on the true discount rate in the warrant market. The growth rate at m is quite easily found since no iteration is involved. The difference in time between m and T will be less the less m is, but no general relation between the size of m and the difference in growth rates at m and T can be obtained.

On first thought one would expect that comparing two warrants, the one with the higher σ_1 should also have the higher σ. The higher the observed risk on the stock, the higher one would expect the risk estimated from the warrant to be. But as can now be seen, since different warrants have different $\frac{Ps}{a}$ and σ_1 in order to have the same growth rates different "time horizons" for different warrants will be necessary. The results show that this is indeed the case.

The fact that the holding period and the "time horizon" are distinct raises problems as to the interpretation of P_e since the price one would be willing to pay for leverage must surely go up as the holding period increases. Since little can be found out about the length of holding periods, it seems impossible to obtain values of P_e per unit of time. The primary purpose of estimating P_e is to obtain a measure of preferences for risk, and this is not affected by the length of the holding period. For instance, if P_e is positive this will always mean that the market wants to take on risk. The influence of a given value of P_e on the price paid for a warrant can be easily obtained. Whether P_e is positive or negative the absolute value of the percentage difference it makes in the price of the warrant will be the same. For a given t, the higher $\frac{Ps}{a}$ is the less the effect of a given P_e on the price of the warrant, and for given $\frac{Ps}{a}$ the greater t is the less the effect.

8. The Data

The data used for this analysis are from two periods, 1923 to 1932, and 1953 to 1959. These two periods were chosen primarily for their inherent interest; the first because it is the classic boom and bust period of modern times, and the second because of its nearness to the present. Within these two periods the samples to be used for analysis were chosen with three considerations in mind.

The first was that time periods be chosen so that the parameters can be assumed to be relatively stable. If this were not the case the estimates of the parameters would be worthless since they would be changing within the period itself. The most likely time periods to choose are those during which there are definite price movements in only one direction--a bear or a bull market, but not both lumped together. Thus for those warrants which existed prior to 1929, there were two samples, one from January 1928 to October 1929 and the other from November 1929 to June 1932, the observations being of the closing price of the first day of each month. Periods of time during which there is little movement either way in prices were also avoided, since this means that there is little variation in the independent variable, stock prices, and thus the estimate obtained will be very inexact and quite meaningless. Therefore, the periods chosen are solely those in which there was a reasonably large price movement

in one direction. In order to make the price movements as large as possible with the 1950's data, samples were chosen on the basis of the highs and lows for each individual stock making the exact interval for each sample during the various stock market movements somewhat different. (E. g., some samples are over the period from 3/3/58 to 12/29/58 while others are from 3/24/58 to 12/29/58.)

The second major consideration was that within given time periods the warrants used must have expiration dates which are in the quite distant future. If the expiration dates are close at hand the relevant variance of expected stock prices must decline during the period of observation, and the estimates of variance again will be worthless. This consideration alone rules out the use of all but a few warrants.

The last major consideration was that the volume of sales of any warrant used must be reasonably large--it must have "breadth and depth" if not "resiliency." If the daily volume of sales in a particular warrant is small, the difference in time between closing stock price and closing warrant price is likely to be large. If there also happens to be a substantial movement in the price of the stock during the day, the observed closing prices will be particularly bad proxies for the simultaneous prices which are needed for the analysis. The larger the volume of sales, therefore, the smaller the measurement error. Another reason making volume of sales an important consideration is that the value of the warrant to the marginal investor who buys will generally be closer to the value of the warrant to the marginal investor who sells, and thus the actual price will better reflect their expectations and preferences the greater the volume of sales. If there is a "thin" market the actual price paid is not necessarily a good indication of anyone's expectations and preferences.

The warrants and the time periods for each sample, as well as the exercising prices and expiration dates of the warrants, are listed in Table 1. Also, the table indicates whether weekly or monthly closing prices were used. Altogether ten samples using eight different warrants issued by seven corporations were employed for the late 1920's - early 1930's period. All but one of these warrants had no expiration date. Only two of these samples were from the pre-crash period. The reason for this was that six of the warrants were issued in 1929 making it impossible to obtain enough observations over this pre-crash period.

For the period from 1953 to 1959, twenty samples involving eight different warrants were used. Only three of the eight warrants had no expiration dates. The other five, besides having expiration dates, were not issued until 1956 or 1957 and so could not be used for earlier periods. During the period from 1953 to 1959 there were five major stock market movements. From January to September 1953 prices fell, from September 1953 to June 1955 prices rose, from January 1957 to August 1957 prices rose, from August 1957 to March 1958 prices fell, and from March 1958 into 1959 prices rose. All of the samples used take account of these movements, but, as mentioned above, the exact periods used for each sample depended upon the highs and lows for the individual stocks and so do not correspond exactly with movements in the stock price indices. The period from July 1955 to January 1957 was remarkable in the stability of stock prices and so no data from this period were used.

The 1928 and 1932 data consist of the daily close for the first trading day of each month obtained from the Journal of Commerce. However, the weekly stock price data needed to test for lognormality, serial correlation, etc., consist of the average of the high and low for the last trading day of each week obtained from the Commercial and Financial Chronicle.

For the 1953 to 1959 period the data consist of the closing prices for each week obtained from Barrons. The monthly data consist of the closing prices for the last complete week of each month.

Before turning to a discussion of the estimates of expectations and preferences obtained through warrant prices, it is first necessary to discuss the results of the various tests on the changes in stock prices. These tests consist of tests for normality or lognormality and serial correlation of changes in stock prices, and the increase in the variance of changes as the time between changes increases.

The data for these tests cover somewhat different periods of time than the warrant-stock data.[1] This resulted in seventeen samples of stock prices for testing purposes. For each of these seventeen samples the changes in stock prices over one, two, four, and sometimes eight and sixteen weeks were tested, the data for the two week changes consisting of the price in week one, price in week three, price in week five, etc. This means that the sample size for changes over two weeks is one-half that for weekly changes. Similarly the sample size for changes over four weeks is one-fourth that for weekly changes, and so on.

9. Normality or Lognormality

One method of testing for normality is to test for skewness and kurtosis. The test criterion for skewness ($\sqrt{b_1}$) is the third moment about the mean divided by the second moment about the mean to the three halves power. For small samples (less than 200) the test criterion for kurtosis (a) is the ratio of the mean deviation $\frac{1}{u}\sum_{i=1}^{u} |x_i - \bar{x}|$ to the standard deviation. A discussion of these test criteria and tables of significance are in Geary and Pearson's paper.[2]

To test for normality the variable to be tested is the change in price of the stock, $x_t - x_{t-1}$. To test for lognormality the variable to be tested is the natural log of the percentage change in price of the stock, $\ln \frac{x_t}{x_{t-1}}$. Weekly changes, and changes over two and four weeks were tested. The results of these tests are given in Table 2. A 1% level of significance was used for these tests. As can be seen the results of these tests are not particularly encouraging. Out of a total of 51 samples tested for skewness 20 samples can be assumed to be skewed too much to be

1. Two warrants were used for the pre-crash period 1928 to October 1929. The tests for the stocks of these two corporations cover the period from July 1927 to October 1929. For the stocks of the warrants used for the post-crash period, the tests cover the period from November 1929 to June 1932. For the three stocks having undated warrants in 1953 the tests cover the stock prices from July 1952 to January 1959. For the other warrants used during the 1957-1958 period the tests cover their stock prices starting six months before the first warrant data and extending to the end of the warrant data. The reason for the inclusion of stock prices before the start of the warrant data for testing purposes is that the behavior of these stock prices will most likely influence expectations at least at the beginning of the period.

2. R. C. Geary and E. S. Pearson, Tests for Normality, Biometrika Office, University College, (London, 1938).

Table 1.

Warrant	Period	Weekly or Monthly Data	Exercising Price	Expiration Date
American & Foreign Power	3/28 - 10/29	Monthly	$25.00	infinite
"	11/29 - 6/32	"	"	"
American Investors	"	"	20.00	"
Commonwealth & Southern	"	"	30.00	"
Electric Power & Light	1/28 - 10/29	"	25.00	"
"	11/29 - 6/32	"	"	"
Niagara Hudson Power Class A	"	"	35.00	1/1/44
" Class B	"	"	14.28	infinite
Tri-Continental	1/30 - 6/32	"	22.50	"
United Corporation	11/29 - 6/32	"	27.50	"
Alleghany Corporation	1/5/53 - 9/21/53	Weekly	3.75	"
"	9/28/53 - 6/27/55	"	"	"
"	9/2/57 - 3/31/58	"	"	"
"	4/7/58 - 12/29/58	"	"	"
"	1/53 - 12/58	Monthly	"	"
Atlas Corporation	1/5/53 - 9/21/53	Weekly	25.00	"
"	9/28/53 - 6/27/55	"	"	"

Warrant	Period	Weekly or Monthly Data	Exercising Price	Expiration Date
Atlas Corporation	2/4/57 - 12/30/57	Weekly	$ 6.25	infinite
"	1/53 - 12/58	Monthly	___a	"
Molybdenum	3/3/58 - 12/29/58	Weekly	30.00	10/18/63
Northspan Uranium	6/10/57 - 2/3/58	"	3.00	12/31/66
Sperry Rand	3/24/58 - 12/29/58	"	25.00	9/16/63
Symington Wayne	3/24/58 - 12/29/58	"	10.00	5/31/63
Tri-Continental	1/5/53 - 9/21/53	"	17.76	infinite
"	9/28/53 - 6/27/55	"	"	"
"	2/4/57 - 8/26/57	"	"	"
"	10/21/57 - 12/29/58	"	"	"
"	1/53 - 12/58	Monthly	"	"
Van Norman	2/4/57 - 12/30/57	Weekly	16.50	3/1/65
"	1/6/58 - 12/29/58	"	"	"

a. In mid 1956 Atlas Corporation had a four for one stock split. At the same time the old warrants were split four for one. The monthly data used for the whole period from 1953 to 1959 consist of the actual stock and warrant prices up to the stock split and after that time the actual prices paid for both multiplied by four.

lognormally distributed and 29 samples too much to be normally distributed. Out of a total of 37 samples tested for kurtosis 18 samples can be assumed to be either too peaked or too flat to be lognormally distributed and 24 samples are too peaked or flat to be normally distributed. In most cases where there is significant skewness the distribution is skewed to the right. The chances of large weekly (or two or four weekly) rises is greater than the chances of large falls. In almost all cases where the criterion for kurtosis is significant, the distributions are too peaked. Too many changes are very small compared with a normal or lognormal distribution. This last result immediately rules out the possibility of using rectangular distributions. The

Table 2.

Number of samples having skewness or kurtosis significantly different from normality or lognormality

Length of Change in Weeks	Number of Samples	$\sqrt{b_1}$ (skewness) Lognormal	Normal	Number of Samples	a (kurtosis) Lognormal	Normal
1928 through 1932						
1	9	5	6	9	6	8
2	9	4	8	9	3	6
4	9	1	1	0	-	-
Total	27	10	15	18	9	14
1953 through 1958						
1	8	5	6	8	5	5
2	8	3	5	8	2	3
4	8	2	3	3	1	2
Total	24	10	14	19	8	10
All Samples						
Total	51	20	29	37	17	24

ideal distribution would seem to be one which is slightly more peaked than a normal distribution and is slightly more skewed than a lognormal distribution. A distribution with these characteristics is not immediately obvious. But even if such a distribution were found, it also would not fit all the various

samples since a few are skewed to the left and some are relatively less peaked. And some of the samples which are not significantly different from normality or lognormality would undoubtedly be significantly different from any new kind of distribution. We can conclude that no one distribution would fit all the samples.

The results presented in Table 7 indicate that the expectations of investors in the warrant market extend quite far into the future. This would imply that it is really the distribution of changes over periods of longer than a week or two which are important. Table 2 shows that for changes over four week periods both distributions are better assumptions than they are for changes over shorter periods. This result is not caused merely by the decrease in sample size, making it harder to obtain significant values of the criteria. The point estimates of the criteria are in general better than they are for the point estimates of the criteria for weekly and two weekly changes. Some work done on changes over eight and sixteen weeks indicates that the point estimates of the criteria over these longer periods are also closer to expectation than those for the very short periods.

With respect to both skewness and kurtosis the assumption of lognormality is definitely superior to normality; a lognormal distribution is a good assumption more than half the time while a normal distribution is a good assumption less than half the time. If we are interested in the relative advantage of using lognormal distributions over normal distributions, for 38 out of 51 samples a lognormal distribution is a better assumption with respect to skewness and for 32 out of 37 samples a better assumption with respect to kurtosis. Out of the 37 samples which were tested for both skewness and kurtosis a lognormal distribution was better with respect to both criteria in 22 cases, and a normal distribution was better for both only 3 times. In 12 cases a lognormal distribution was better with respect to one criterion while a normal distribution was better with respect to the other. The assumption of lognormality is definitely the better assumption.

10. Serial Correlation

The results of the tests for serial correlation were much more favorable than the above tests. Each sample was tested for serial correlation of price changes over one, two, four, eight, and sixteen week periods. For large samples the test used for the null hypothesis is based on the approximations of R. L. Anderson.[1] The mean and variance of the serial correlation coefficient are approximately $\frac{-1}{n-1}$ and $\frac{n-2}{(n-1)^2}$ respectively. For small samples (under 60) tables of significance for the non-circular serial correlation coefficient have been developed by T. W. Anderson.[2] Since in all the samples tested there were substantial trends, the circular serial correlation coefficient is not usable. A 1% level of significance was used for these tests.

There were 17 samples each tested for changes over the five different length periods for a total of 85 samples. Out of these 85 samples only one has any significant serial correlation (at the one per cent level).

1. R. L. Anderson, "Distribution of the Serial Correlation Coefficient," Annals of Mathematical Statistics, vol. 13, (1942), pp. 1-13.

2. T. W. Anderson, "On the Theory of Testing Serial Correlation," Skandinavisk Aktuarietidskrift, vol. 31, (1948), p. 114.

This sample is of changes over one week and has large negative serial correlation (R = -.35). The negative coefficient implies that if the price moves up one week it will most likely move down the next week. The size of the coefficients classified by sample size are shown in Table 3. Of the 85 samples 42 had positive coefficients and 43 had negative coefficients, thus, there is no preponderance in either direction.

Table 3.

Number of samples having absolute value of R within the ranges:

	0-.1	.1-.2	.2-.3	.3 up
Sample Size Over 100	8	6	4	1
50 to 100	7	7	3	-
25 to 50	7	7	3	-
Under 25	8	8	4	11

The conclusion drawn from the results of the tests for serial correlation is that almost without exception there is no significant correlation of changes in these stock prices over periods through sixteen weeks. This result upholds the assumption that stock prices follow a simple random walk model. However, before accepting a simple random walk model the relation between the variance of changes in stock prices and time must be examined.

11. Variance and Time

It was shown above that the variance of changes in a variable which follows a simple random walk increases proportionately to the length of time of the changes. Therefore, for a simple random walk model to apply to changes in stock prices not only must the changes be uncorrelated, but the variance must, for instance, be twice as great for changes over a two week period as over a one week period. There are certain complications in testing for this relationship.

One possible method is to use the F-test on the ratios of the variances. We want to test whether $2\sigma_1^2 = \sigma_2^2$, $4\sigma_1^2 = \sigma_4^2$, $8\sigma_1^2 = \sigma_8^2$, $16\sigma_1^2 = \sigma_{16}^2$ where σ_1^2 = variance of changes over a one week period, σ_2^2 = variance of changes over a two week period, etc. The ratios of $\frac{t\sigma_1^2}{\sigma_t^2}$ are distributed as F, but only if σ_1^2 and σ_t^2 are inde-

pendent. The assumption of independence is clearly invalid in this case. For this test both σ_1^2 and σ_t^2 are estimates subject to sampling error of true parameter values. If $t\sigma_1^2$ were considered to be a true population value subject to no sampling error, the ratio $\frac{n\sigma_t^2}{t\sigma_1^2}$, where n = sample size -1, is a chi-square variable. The fact that only the one term in this ratio has any sampling error means that this latter test is more stringent (i.e., more likely to obtain significant differences) than the F-test. The assumption that σ_1^2 has no sampling error is, of course, invalid, but since it is estimated in all cases from large samples its error is small, and the chi-square test may be reasonable. One would expect that the bias in using this test when σ_1^2 is subject to error would be in the direction of obtaining significant differences between $t\sigma_1^2$ and σ_t^2 when in actual fact there are no significant differences. Since neither of the two tests are strictly applicable, it seemed better to bias the results in this direction by using the chi-square test. A 1% level of significance was also used for these tests.

The results of these tests in almost every case show that the simple random walk hypothesis can be accepted. Since changes over a week are taken as given there are a total of 68 instead of 85 samples. Of these 68 only four of the variances were significantly different from the values expected. For both of the stocks tested over the period June 1927 to October 1929, the variance of changes over two weeks was significantly different from twice the variance of weekly changes. For one of these two the variance of changes over four weeks was also significantly different from four times the variance of weekly changes. The other instance of significance was of the variance of changes over two weeks for one of the stocks tested over the period June 1952 to January 1959. There were no cases of significance for the variances of changes over eight and sixteen weeks, but the sample sizes were small. This could mean that although there might be no significant differences between $16\sigma_1^2$ and σ_{16}^2, there could be a definite trend toward σ_{16}^2 being greater or less than $16\sigma_1^2$. This possibility did not arise since nine of the σ_{16}^2 were greater than $16\sigma_1^2$ and eight of the σ_{16}^2 were less.

The conclusions drawn from these tests are that a simple random walk model with no barriers seems to be a not unreasonable model for stock prices and that changes in stock prices can usually be assumed to be log-normally distributed.

12. The Results Using Warrants

As explained above, it was originally intended to estimate an expected standard deviation (σ), a price of leverage (P_e), and a measure (k) of the relation between the mean expected stock price and the present price of the stock, but it was found impossible to obtain these estimates. The standard errors of the parameters increased substantially as the iterative process converged indicating the presence of collinearity among the parameters which were significantly different from the initial assumed estimates of them. The next step was to assume the value of one of the parameters and iterate for values of the other two. This was attempted with the variable k assumed equal to one, but the results from using this method showed no improvement over the method which attempts to estimate all three parameters. Because of this it was necessary to assume values of two of the parameters and iterate for the value of the third. This was done assuming

always that $k = 1$ and that $P_e = +125$, 0 or $-.25$ and for several samples $+.10$ and $-.10$. This meant that the iterative process had to be done three (sometimes five) times for each sample--one for each assumed value of P_e.

Admittedly the results using the last method will leave something to be desired in comparison with using the first method in the best of all possible worlds. Just what has been given up and what problems have arisen using the last method?

First k is assumed equal to one at all times. In Section 3 it was seen that there are two reasons why investors might pay for leverage: (1) they are risk lovers and (2) they are risk averters, but they think that the price of the stock will rise. If we knew the true value of k, (2) would be accounted for and P_e would measure solely (1). The estimate of P_e would then be an accurate measure of the preferences toward risk of the marginal investor. Not being able to estimate k then means that P_e will no longer measure just preferences toward risk. Statistically, however, there is no reason to assume that this will affect solely the estimates of P_e and not the estimates of σ. For instance, if k is actually greater than one this assumption of $k = 1$ will tend to raise the estimates of both P_e and σ but by how much is indefinite. The effects of assuming $k = 1$ would be particularly bad for growth stocks since for these stocks both the buyer and seller could easily expect the future mean price to be higher than the present price. From Table 4 it will be seen that with one exception the point estimates of k lie quite close to one, and the standard errors in these cases are sometimes quite small. In the one case of $k = .75$ on the other hand, the standard error is very large. These results seem to indicate that the true values of k are usually quite close to one so that the necessary assumption of $k = 1$ should not affect the results much. The last method consisting of iterating for only one coefficient assuming values for the other two, could have been used to obtain maximum likelihood estimates of both σ and P_e, but the computing time would have been excessive.[1] For this reason no attempt was made to obtain these estimates.

In Table 5 the results of iterating for σ for different assumed values of P_e are detailed. The values of σ with $P_e = 0$ are shown as are the values of σ with P_e equal to .25, 0, or $-.25$ whichever yields the highest value of the likelihood function. The standard errors of estimate for both pairs of values are also shown. Knowing the standard error of estimate it is easy to see how good the fit is by obtaining the r^2 for each sample since $r^2 = 1 - \frac{\sigma^2_{v-Pw}}{\sigma^2_{Pw}}$ where σ^2_{v-Pw} = the residual sum of squares and σ^2_{Pw} = the variance of Pw. The r^2 for each

1. For each assumed value of P_e and final estimate of σ the value of the likelihood function can be obtained. (The standard error of estimate is an easily obtained proxy for the likelihood function; when it is minimized the likelihood function is maximized.) Thus, by assuming enough different values of P_e and for each of them iterating for σ, maximum likelihood estimates can be obtained to any degree of accuracy. However, for each assumed value of P_e usually three and sometimes four iterations were necessary to obtain the value of σ, and the computing time for each observation per iteration was approximately six seconds making the necessary computing time unreasonably long.

Table 4.

Results Estimating All Three Parameters

Warrant and Time Period	P_e	σ	k	Standard Error of Estimate
Alleghany 9/2/57 to 3/31/58	-.669	1.934	1.040	.158
	(4.670)	(1.903)	(.344)	
Alleghany 4/7/58 to 12/29/58	-1.192	2.050	1.029	.192
	(5.351)	(1.916)	(.251)	
Alleghany 1953 thru 1958 (monthly)	1.340	1.459	.750	.214
	(5.754)	(1.414)	(.849)	
Tri-Continental 2/4/57 to 8/26/57	.006	.083	.995	.130
	(.134)	(1.435)	(.077)	
Tri-Continental 10/21/57 to 12/29/58	-.307	.472	1.149	.505
	(.307)	(.226)	(.118)	
Tri-Continental 1953-1958 (monthly)	-.404	.719	1.112	.698
	(.239)	(.447)	(.071)	

(The figures in parentheses are the standard errors of the estimates.)

---

sample is shown in Table 5 next to the figures for the standard error of estimate. With several exceptions the r^2's are extremely high, the exceptions always occurring where the range of Ps was very small. The differences between the standard error of estimate for the "best" P_e and for $P_e = 0$ are in most cases small. By just assuming $P_e = 0$ one would not lose much in the way of a worse fit in most cases. For the six samples for which results were obtained by iterating for all three parameters, the standard errors of estimate in four cases are smaller than, and in two cases are equal to, the standard errors shown in Table 5 for these six samples. In the two cases where they are equal the estimated values of k are the closest to the assumed value of one. As would be expected, iterating for only one parameter leads to worse fits than iterating for all three parameters.

The values of P_e shown in Table 5 lead to some interesting conclusions. The actual values themselves do not mean much, but the direction in which they tend does. As long as the likelihood function has only one maximum, if the best value of P_e is -.25 then it is definite that the true value of P_e is negative, although whether greater or less than -.25 is not known. It has not been proved that the likelihood function in this case has only one maximum, but the results of the large total number of iterations made give the impression that there is only one maximum. It might be expected that when prices are rising people are more willing to take on risk and when falling more unwilling to take it on. Thus it would seem reasonable to expect P_e to be higher in upswings than in downswings. This expectation would be strengthened by the fact that k also would more likely be higher in upswings than in downswings, and this would increase the effect

Table 5.

Results Estimating Only One Parameter

Warrant	Time Period	σ for $P_e=0$	σ for "Best" P_e	"Best" P_e	Standard Error of Estimate for "Best" P_e	r^2	Standard Error of Estimate for $P_e=0$
American & Foreign Power	3/28 thru 10/29	1.005 (.028)	1.005 (.028)	0	2.043	.998	2.043
American & Foreign Power	11/29 thru 6/32	1.131 (.022)	1.404 (.018)	-.25	2.460	.991	2.460
American Investors	"	1.623 (.060)	1.781 (.058)	-.25	.809	.95	.822
Commonwealth & Southern	"	1.253 (.042)	1.397 (.041)	-.25	.361	.96	.402
Electric Power & Light	1/28 thru 10/29	1.015 (.022)	1.298 (.018)	-.25	1.505	.99	1.753
"	11/29 thru 6/32	.938 (.021)	1.250 (.017)	-.25	1.629	.993	1.924
Niagara Hudson Power Class A	"	1.224 (.036)	1.366 (.035)	-.25	.354	.96	.384
Niagara Hudson Power Class B	"	1.553 (.049)	1.351 (.046)	+.25	.510	.98	.543
Tri-Continental	"	1.489 (.049)	1.636 (.047)	-.25	.800	.88	.839
United Corp.	"	1.219 (.021)	1.414 (.020)	-.25	1.332	.98	1.566

Table 5 (Continued)

Warrant	Time Period	σ for $P_e=0$	σ for "Best" P_e	"Best" P_e	Standard Error of Estimate for "Best" P_e	r^2	Standard Error of Estimate for $P_e=0$
Alleghany Corp.	1/5/53 thru 9/21/53	1.715 (.147)	1.546 (.140)	+.25	.200	.33	.200
"	9/28/53 thru 6/27/55	1.483 (.082)	1.679 (.076)	-.25	.170	.98	.197
"	9/2/57 thru 3/31/58	1.620 (.145)	1.700 (.140)	-.10	.158	.98	.158
"	4/7/58 thru 12/29/58	1.373 (.109)	1.601 (.097)	-.25	.192	.98	.200
"	1/53 thru 12/58 (monthly)	1.476 (.086)	1.687 (.078)	-.25	.247	.98	.268
Atlas Corp.	1/5/53 thru 9/21/53	.385 (.017)	.385 (.017)	0	.235	.78	.235
"	9/28/53 thru 6/27/55	.296 (.016)	.296 (.016)	0	.385	.996	.385
"	2/4/57 thru 12/30/57	.880 (.071)	1.135 (.050)	-.25	.212	.94	.263
"	1/53 thru 12/58 (monthly)	.614 (.012)	.909 (.010)	-.25	3.464	.59	3.705
Molybdenum Corp.	3/3/58 thru 12/29/58	1.598 (.019)	1.345 (.019)	+.25	1.669	.81	1.682

Table 5 (Continued)

Table 5 (Continued)

Warrant	Time Period	σ for $P_e=0$	σ for "Best" P_e	"Best" P_e	Standard Error of Estimate for "Best" P_e	r^2	Standard Error of Estimate for $P_e=0$
Northspan Uranium	6/10/57 thru 2/3/58	1.567 (.159)	1.768 (.146)	-.25	.228	.97	.239
Sperry Rand	3/24/58 thru 12/29/58	1.089 (.021)	.887 (.021)	+.25	.766	.87	.863
Symington Wayne	"	1.354 (.048)	1.120 (.052)	+.25	.448	.94	.507
Tri-Continental	1/5/53 thru 9/21/53	.639 (.026)	.847 (.025)	-.25	.182	.78	.224
"	9/28/53 thru 6/27/55	.451 (.016)	.741 (.013)	-.25	.728	.91	1.031
"	2/4/57 thru 8/26/57	.080 (.568)	.080 (.568)	0	.141	.996	.141
"	10/21/57 thru 12/29/58	.353 (.064)	.353 (.064)	0	.651	.98	.651
"	1/53 thru 12/58 (monthly)	.425 (.021)	.782 (.014)	-.25	.833	.97	.843
Van Norman	2/4/57 thru 12/30/57	1.275 (.045)	1.425 (.043)	-.25	.665	-.04[1]	.701
"	1/6/58 thru 12/29/58	1.551 (.044)	1.384 (.044)	+.25	.854	.64	.873

(The figures in parentheses are the standard errors of the estimates).

1. The range of the price of the stock was so small for this sample that the mean of Pw is a relatively good estimator of Pw over the period. With a linear regression it would be impossible to obtain a standard error higher than the standard deviation of Pw , but with a non linear regression this is possible. The non linear regression leads to a worse explanation of Pw in this case than simply using the mean Pw .

on P_e if k were assumed equal to one. The results shown in Table 6 tend to support this conclusion. During upswings P_e is positive or zero better than half the time, while in downswings P_e is overwhelmingly negative.[1]

Table 6.

	No. of Samples for which $P_e = +.25$	No. of Samples for which $P_e = -.25$	No. of Samples for which $P_e = 0$
Upswings	4	6	4
Downswings	2	13	1

Results of the comparison of the standard deviations estimated from the warrants with weekly observed standard deviations are shown in Table 7. The warrants are listed first by time periods and within each period by decreasing value of $\frac{Ps}{a}$. Column 1 shows $\frac{Ps}{a}$ where Ps is the mean price of the stock during the period. Column 2 shows the standard deviation estimated from the warrant, and Column 3 shows the observed standard deviation of weekly changes in stock prices. Column 4 is the number of years "time horizon" (m) implied by the ratio $\frac{\sigma^2}{52\sigma_1^2}$. As was seen in Section 7 there is no reason to expect each warrant within any one period to have the same horizon because of the influence of $\frac{Ps}{a}$ on the percentage growth rate of the warrant. Even if this effect were neglected the "time horizons" need not be the same; one would then expect just that σ would be higher the higher σ_1 is.

In Section 7 it was also seen that the estimated σ and its implied "time horizon" set an upper limit on the true value of the discount rate in the warrant market and also s ets a lower limit on the actual "time horizon" for each warrant. Given σ and σ_1 the percentage rate of growth at time m can be easily found by obtaining $\left[\frac{1}{E[Pw]^*}\right]\left[\frac{dE[Pw]^*}{dt}\right]$. The value of this rate of growth (i) is shown in column 5 for $P_e = 0$. If $P_e \neq 0$

1. One other interesting result shown in Table 5 should be mentioned. There seems to be an approximately linear relation between P_e and σ for any warrant. When P_e is .25, the estimated σ is decreased about $\Delta\sigma$ from when P_e is assumed 0, and when P_e is -.25, the estimated σ is increased by about the same $\Delta\sigma$. Not only is this relation approximately linear, but it is in some cases close to one--$\Delta\sigma$ would be approximately .25. For most samples, however, it seems to be closer to 4/5 than to 1; if P_e is $\pm$ .25, $\Delta\sigma$ is about $\pm$ .2. This relation means that knowing σ when $P_e = 0$, it is possible to approximate σ for any value of P_e without going through the whole iterative technique.

Table 7.

Warrant	Time Period	Col. 1 P_s a	Col. 2 for "Best" P_e	Col. 3 Weekly Observed	Col. 4 m in years	Col. 5 i for $P_e = 0$	Col. 6 i for "Best" P_e
American & Foreign Power	3/28-10/29	3.17	1.005	.0834	3	2.6%	2.6%
Electric Power & Light	1/28-10/29	2.04	1.298	.0667	7	2.6%	2.9%
Electric Power & Light	11/29-6/32	1.77	1.250	.0995	3	7.4%	8.2%
American & Foreign Power	11/29-6/32	1.79	1.404	.1265	2.5	9.7%	10.6%
United Corp.	11/29-6/32	.92	1.414	.0836	5.5	8.2%	9.4%
Niagara Hudson Power B	11/29-6/32	.85	1.351	.0783	6	8.8%	7.7%
American Investors	11/29-6/32	.42	1.781	.1462	3	24.9%	28.0%
Tri-Continental	1/30-6/32	.41	1.636	.0943	6	15.4%	18.1%
Niagara Hudson Power A	11/29-6/32	.35	1.366	.0783	6	23.9%	36.4%
Commonwealth & Southern	11/29-6/32	.33	1.397	.0695	8	20.9%	36.7%
Alleghany	4/7/58-12/29/58	1.89	1.601	.0586	14.5	1.5%	1.7%
Tri-Continental	10/21/57-12/29/58	1.89	.353	.0206	5.5	.4%	.4%
Alleghany	1/53-12/58	1.78	1.687	.0541	14	1.3%	1.3%
Tri-Continental	2/4/57-8/26/57	1.74	.080	.0227	.25	.0%	.0%

Table 7 (Continued)

Warrant	Time Period	Col. 1 $\frac{Ps}{a}$	Col. 2 for "Best" P_e	Col. 3 Weekly Observed	Col. 4 m in years	Col. 5 i for $P_e=0$	Col. 6 i for "Best" P_e
Northspan	6/10/57-2/3/58	1.61	1.768	.0672	13.5	2.0%	2.1%
Atlas	9/28/53-6/27/55	1.46	.296	.0230	3	2.0%	2.0%
Atlas	2/4/57-12/30/57	1.45	1.135	.0395	16	1.8%	1.8%
Tri-Continental	1/53-12/58	1.45	.782	.0306	13	1.9%	2.4%
Alleghany	9/2/57-3/31/58	1.44	1.700	.0485	24	1.2%	1.2%
Atlas	1/53-12/58	1.43	.909	.0344	13.5	2.2%	2.7%
Alleghany	9/28/53-6/27/55	1.42	1.679	.0573	17	1.7%	1.9%
Tri-Continental	9/28/53-6/27/55	1.22	.741	.0510	4	8.3%	11.2%
Atlas	1/5/53-9/21/55	1.18	.385	.0110	13.5	1.1%	1.1%
Alleghany	1/5/53-9/21/55	1.15	1.546	.0570	14	2.5%	2.3%
Symington Wayne	3/24/58-12/29/58	1.07	1.120	.0491	10	4.1%	3.5%
Tri-Continental	1/5/53-9/21/53	.90	.847	.0284	17	2.7%	3.3%
Molybdenum	3/3/58-12/29/58	.88	1.345	.0615	9	5.4%	4.7%
Sperry Rand	3/24/58-12/29/58	.80	.887	.0357	12	4.6%	3.8%
Van Norman	1/6/58-12/29/58	.52	1.384	.0665	8	9.2%	7.9%
Van Norman	2/4/57-12/30/57	.50	1.425	.0622	10	7.7%	9.2%

this will not change the dollar rate of growth, $\frac{dE[Pw]^*}{dt}$, of the expected exercising value, but it does change the present price. If $E[Pw]'_1$ is the present price with $P_e \neq 0$, then the percentage rate of growth is $\frac{dE[Pw]^*}{dt} \cdot \frac{1}{E[Pw]'_1}$. The values of this at m years are given in column 6. For the late 1920's and the 1950's either of these measures gives low discount rates, but for the early 1930's the discount rates are quite high. It should be remembered that these discount rates are upper limits to the true value of the discount rate which then must be quite small in all periods except the early 1930's. Within all three periods the estimates of the discount rates seem to be quite consistent. When the large number of places where errors can arise is remembered, the consistency of the results within periods seem quite extraordinary. The estimates of the σ's and P_e's can be quite poor since they are not maximum likelihood estimates; the estimates of the σ_1's can be poor because of lack of knowledge as to the relevant periods; and the use of the mean Ps for the ratio $\frac{Ps}{a}$ may lead to errors.

13. Conclusions

The primary purpose of this study was to obtain empirical estimates of certain investors' expectations and preferences. In order to obtain these estimates a theory of common stock warrant prices was formulated which showed that the price of any warrant depends upon the market's or marginal investor's expectations as to the future mean price of the stock, the expected risk of the stock as measured by the standard deviation of expected stock prices, and his preferences toward risk. Since tests did not disprove that actual stock price changes are usually lognormally distributed and follow a simple random walk, it was assumed that expected stock prices have these characteristics. If expected stock prices follow a simple random walk, the amount of risk increases as one looks further into the future, and thus the price of a warrant will increase as the market looks further into the future even though the expected mean stock price remains the same.

Although it was found impossible to estimate all the variables which are of interest and are theoretically obtainable, the results actually obtained readily yield some interesting conclusions.

It was found that warrant prices can be explained extremely well through the use of the model developed here. This result means that warrant prices do depend upon the risk or standard deviation of stock prices as the theory suggests.

One would expect that the estimates of expected standard deviations (σ) would be correlated with the weekly observed standard deviations (σ_1) since previous performance of a stock should affect expectations. Other things equal, the higher σ_1 is, the higher σ should be. There were too few samples for the period from 1928 to 1932 to allow a test for this relation, but for the 1950's data the σ's and σ_1's were correlated. For this latter period r^2 was .761, which is significant for a sample size of 20. There does seem to be a significant positive relation between σ and σ_1. The expected risk of a stock seems to be higher the higher the actual risk of the stock, and an increase in the actual risk of a stock should lead then to an increase in the expected risk of the stock. Since the measures of actual risk were obtained from observations starting only six months or less in the past, the high correlation between σ_1 and σ implies that expected risk is highly influenced by the very recent past.

But, as we have seen, other factors may enter in so that the σ's

will not solely depend upon σ_1. The ratio of the mean price of the stock to the exercising price $\frac{Ps}{a}$ will influence the expected percentage rate of growth of the exercising value of the warrant (i), and therefore, may affect the value of σ. For a given σ_1 the higher either $\frac{Ps}{a}$ or σ is the lower will i be. Thus for a warrant with a high $\frac{Ps}{a}$ in order to keep i as high as it would be for other warrants with lower $\frac{Ps}{a}$, σ must be lower. One would thus expect that for a given σ_1, the higher $\frac{Ps}{a}$ is the lower will σ be in order to keep i close to the same value for different warrants. Also if the true value of k is different from unity the relation between σ and σ_1 will be affected. The greater the true value of k the higher will be the estimate of σ. There is no way of directly obtaining values for k, but there is an indirect method of approximating its effect. The most powerful factor affecting k should be the dividend yield or ratio of annual dividends to the mean stock price $\frac{D}{Ps}$. If a corporation pays no dividends, investors in its stock must be expecting a capital gain, or they would not be willing to pay the current market price for the stock. Thus if $\frac{D}{Ps}$ is zero, k should be greater than one, and in general the higher $\frac{D}{Ps}$ is the lower k should be. One would expect that the higher $\frac{D}{Ps}$ is the lower will σ be.

The effects of σ_1, $\frac{Ps}{a}$, and $\frac{D}{Ps}$ on the estimates of the σ's for the 1950's were tested by means of a multiple regression. The estimated relation with the standard errors of the coefficients below in parentheses is,

$$\sigma = +.649 + 19.440\,\sigma_1 - .103\,\frac{Ps}{a} - 7.845\,\frac{D}{Ps}$$
$$(4.18) \quad (.154) \quad (2.69)$$

and r^2 is .850, which is significant. Both σ_1 and $\frac{D}{Ps}$ are significantly related to σ, but $\frac{Ps}{a}$ does not significantly affect σ. A rise in σ_1 of .01 will result in a rise in of about .19, and this compares quite well with the average of the $\frac{\sigma}{\sigma_1}$ ratios. The effect of the dividend yield on σ although significant seems to be rather small as might be expected since $\frac{D}{Ps}$ is an imperfect substitute for k. If there is a reasonable relation between $\frac{D}{Ps}$ and k, we can conclude that both k and σ_1 affect the estimates of σ.

The fact that $\frac{Ps}{a}$ does not significantly affect σ implies that the market's expectations as to risk are not significantly affected by the percentage growth rate of the warrant (i). One would expect then that $\frac{Ps}{a}$ would significantly affect the estimates of i. The effects of $\frac{Ps}{a}$, and σ, σ_1, and $\frac{D}{Ps}$ on the estimates of the i's for the 1950's were tested by means of another regression. Of course the higher $\frac{Ps}{a}$ is the

lower should i be. Since for a given σ_1 the higher σ is the longer the "time horizon," and the lower i should be. And for a given σ the higher σ_1 is the shorter the "time horizon," and the higher i should be. $\frac{D}{Ps}$ is used as a proxy for the true value of k such that when k is high then $\frac{D}{Ps}$ is likely to be low. The higher k is the higher i would be, so the lower $\frac{D}{Ps}$ is the higher i would be. The i's used are those derived on the assumption that $P_e = 0$ (Column 5 of Table 7). The estimated relation is,

$$i = +.703 - .443\ \frac{Ps}{a} - .089\sigma + .102\sigma_1 - .241\ \frac{D}{Ps}$$
$$(.043) \quad (.069) \quad (4.84) \quad (.938)$$

and r^2 is .545, which is significant. The only variable which has anywhere near a significant effect on i is $\frac{Ps}{a}$. The effect of $\frac{Ps}{a}$ is very strong. For instance a rise of .1 in $\frac{Ps}{a}$ will result in a decrease of .04 or 4% in i. The conclusion must be that $\frac{Ps}{a}$ affects mainly the i's and not the σ's. When stock prices are high it is difficult to get a reasonable return by investing in warrants. In fact for Tri-Continental warrants over several periods the i's are virtually 0%. This helps to explain why Tri-Continental warrants were being exercised in large numbers during these periods. With the price of the stock high the price for the warrant which can be obtained by selling it in the market is very close to the value of the warrant if it is exercised. When transactions costs are taken into account little if anything is to be gained by selling instead of exercising the warrant, and the growth rate is so low that holding the warrant will lead to little gain. It must be remembered that the i's are estimated on the assumption that the stocks are a fair bet--the expected mean price of the stock is equal to the present price. A low i implies then that any one who holds the warrant and thinks the stock is a fair bet must be a risk lover. But an investor may hold a warrant with a low i not because he is a risk lover but because he believes the mean expected price of the stock is greater than its present price. As long as there are some of these "bullish" investors not all the Tri-Continental warrants (or other warrants with low i's) will be exercised.

The fact that the estimated growth rates are so very low for almost all warrants during the late 1920's and the 1950's leads to some interesting questions. The low i's imply that either the warrant investors were risk lovers who thought stocks a fair bet, or they were optimistic as to future rises in the prices of the stocks. The few estimates of k which were obtained seem to indicate that the market did not expect the stocks to rise much. In general then it would seem that the market consisted mainly of risk lovers who would be willing to pay a positive price for leverage. Since warrants are such speculative and risky assets this implication is not surprising. But options in general have two distinct purposes. They can be used for insurance (hedging), or they can be used for their leverage to take on more risk. Warrants, however, are very inefficient assets for hedging purposes; other types of options being much less costly. One would expect, then, that most investors in the warrant market would be risk lovers. The low growth rates during these periods is not surprising.

However, the estimated values of the price of leverage seem to contradict the conclusion that predominantly risk lovers are in the warrant market. As was seen P_e is almost always negative in downswings, but in

upswings it is only positive about half the time. This result suggests that investors in the warrant market do not like to take on risk, but in an upswing they may dislike risk less--sometimes enough less so that they are willing to pay to take it on. Furthermore, for the late 1920's and 1950's P_e was positive for only five samples and negative for 12, and only 4 of the 12 negative P_e's were during downswings. The values of P_e then give some support to the conclusion that investors in the warrant market are generally risk averters. However, this result certainly cannot be considered clear-cut since the P_e's are not overwhelmingly in one direction or another. The estimated growth rates, on the other hand, are all very low giving clear evidence the investors in the warrant market during these periods were risk lovers.

In contrast to the 1920 and 1950 results, the results for the early 1930's seem to be completely consistent. The estimated growth rates are high and the P_e's are almost all negative clearly indicating that investors in the warrant market did not like to take on risk. They had to be given greater inducements in the form of higher rates of growth in order to induce them to hold such speculative assets as warrants. This seems an entirely reasonable attitude on their part for such a period as the early 1930's.

The results of this study seem to lead to conclusions which are consistent with a priori expectations. Investors in the warrant market have expectations which are clearly related to past events, and the very recent past seems to affect expectations very strongly. The overwhelming evidence of the i's suggest that generally they are willing to accept quite low average returns in exchange for obtaining more risk, but in periods when any asset is very risky they will not pay to take on risk.

Since the method of obtaining measures of expectations and preferences toward risk used here has proved reasonably successful, it would seem quite worthwhile to use it to obtain similar measures for investors in convertible bonds. One might expect that these investors have preferences which may be closer to those of the majority of investors than the preferences of investors in warrants, and thus the estimates obtained would have a more general usefulness. Also these investors may have expectations which are based more heavily upon the less recent past than those of investors in the warrant market.

APPENDIX A

For an investor who thinks future stock prices are lognormally distributed[1] with a mean of $\ln x$ of μ_x and a variance of $\ln x$ of σ_x^2,

$$f(x) = \frac{1}{x\sigma_x\sqrt{2\pi}}\, e^{-1/2\left[\frac{\ln x-\mu_x}{\sigma_x}\right]^2} \qquad 0 \leq x \leq \infty \qquad \text{(A.1)}$$

where

$$\frac{1}{\sigma_x\sqrt{2\pi}}\int_0^\infty \frac{1}{x}\, e^{-1/2\left[\frac{\ln x-\mu_x}{\sigma_x}\right]^2} dx = 1 . \qquad \text{(A.2)}$$

This is easily seen by putting $\ln x = x'$, $x = e^{x'}$, and $dx = e^{x'}dx'$. When $x = 0$, $x' = -\infty$.

---

1. See Aitchison and Brown, op. cit., for an excellent discussion of the properties of the lognormal distribution.

Then,

$$\frac{1}{\sigma_x\sqrt{2\pi}} \int_0^\infty \frac{1}{x} e^{-1/2\left[\frac{\ln x - \mu_x}{\sigma_x}\right]^2} dx \quad \frac{1}{\sigma_x\sqrt{2\pi}} \int_{-\infty}^\infty e^{-1/2\left[\frac{x'-\mu_x}{\sigma_x}\right]^2} dx' = 1. \tag{A.3}$$

The expected value E(x) of a lognormally distributed variate is given by:

$$E[x] = \int_0^\infty \frac{x}{x\sigma_x\sqrt{2\pi}} e^{-1/2\left[\frac{\ln x-\mu_x}{\sigma_x}\right]^2} dx = \int_{-\infty}^\infty \frac{e^{x'}}{\sigma_x\sqrt{2\pi}} e^{-1/2\left[\frac{x'-\mu_x}{\sigma_x}\right]^2} dx'$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} \int_{-\infty}^\infty e^{-1/2\left[\frac{x'^2-2\mu_x x'+\mu_x^2-2\sigma_x^2 x'}{\sigma_x^2}\right]} dx'$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} \int_{-\infty}^\infty e^{-1/2\left[\frac{(x' - \mu_x - \sigma_x^2)^2 - \sigma_x^4 - 2\mu_x\sigma_x}{\sigma_x^2}\right]^2} dx'$$

$$= \frac{e^{\mu+1/2\sigma^2}}{\sigma_x\sqrt{2\pi}} \int_{-\infty}^\infty e^{-1/2\left[\frac{x'-\mu_x-\sigma_x^2}{\sigma_x}\right]^2} dx' = e^{\mu_x+1/2\sigma_x^2}. \tag{A.4}$$

The expected value of a warrant is:

$$E[Pw] = \int_a^\infty (x-a)\, f(x)\, dx . \tag{A.5}$$

With a lognormal distribution:

$$E[Pw] = \int_a^\infty \frac{(x-a)}{x\sigma_x\sqrt{2\pi}} e^{-1/2\left[\frac{\ln x-\mu_x}{\sigma_x}\right]^2} dx . \tag{A.6}$$

The assumption is that,

$$E[x] = kPs = e^{\mu_x+1/2\sigma_x^2}$$

$$\mu_x = \ln Ps + \ln k - 1/2\sigma_x^2 \tag{A.7}$$

It is easiest when using a lognormal distribution to define a new variable g, such that $x = Psg$. Then g is any anticipated change in the price of the stock and is also distributed lognormally with a variance of lng equal to σ_g^2 and a mean of lng equal to μ_g. Corresponding to equations (1), the value of the warrant will be,

$$\begin{array}{ll} Psg - a & Psg > a \\ 0 & Psg \leq a \end{array} \qquad \text{(A. 8)}$$

Equation A. 4 becomes,

$$E[g] = e^{\mu_g + 1/2\sigma_g^2} \qquad \text{(A. 9)}$$

and A. 7 becomes,

$$kPs = Ps\, e^{\mu_g + 1/2\sigma_g^2}$$

$$\mu_g = \ln k - 1/2\sigma_g^2 \qquad \text{(A. 10)}$$

Let $\mu_g = \mu$ and $\sigma_g = \sigma$.
Then

$$f(g) = \frac{1}{g\sigma\sqrt{2\pi}}\, e^{-1/2\left[\frac{\ln g - \mu}{\sigma}\right]^2} = \frac{1}{g\sigma\sqrt{2\pi}}\, e^{-1/2\left[\frac{\ln g - \ln k + 1/2\sigma^2}{\sigma}\right]^2} \qquad \text{(A. 11)}$$

Given a particular present price of the stock, the expected value of the warrant to an investor will be,

$$E[Pw;Ps] = \int_{\frac{a}{Ps}}^{\infty} (Psg - a) f(g)\, dg = \int_{\frac{a}{Ps}}^{\infty} \frac{(Psg - a)}{g\sigma\sqrt{2\pi}}\, e^{-1/2\left[\frac{\ln g - \mu}{\sigma}\right]^2} dg\,. \qquad \text{(A. 12)}$$

Let $g' = \ln g$; then $g = e^{g'}$, $dg = e^{g'} dg'$; and when $g = \frac{a}{Ps}$, $g' = \ln a - \ln Ps$.
Then $E[Pw;Ps]$ will be,

$$E[Pw;Ps] = \frac{Ps}{\sigma\sqrt{2\pi}} \int_{\ln a - \ln Ps}^{\infty} e^{g'} e^{-1/2\left[\frac{g' - \mu}{\sigma}\right]^2} dg'$$

$$- \frac{a}{\sigma\sqrt{2\pi}} \int_{\ln a - \ln Ps}^{\infty} e^{-1/2\left[\frac{g' - \mu}{\sigma}\right]^2} dg' \qquad \text{(A. 13)}$$

From equation A. 4 the first term on the right becomes,

$$e^{\mu + 1/2\sigma^2}\, \frac{Ps}{\sigma\sqrt{2\pi}} \int_{\ln a - \ln Ps}^{\infty} e^{-1/2\left[\frac{g' - (\mu + \sigma^2)}{\sigma}\right]^2} dg'$$

Remembering that $e^{\mu + 1/2\sigma^2} = k$, it is seen that both terms are now in the form of constants times normal cumulative frequency functions. Standardizing the variates $\frac{g' - (\mu + \sigma^2)}{\sigma}$ and $\frac{g' - \mu}{\sigma}$, the expected value of the warrant to the investor becomes,

$$E[Pw;Ps] = \frac{kPs}{\sqrt{2\pi}} \int_{-1/2\sigma - \frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2} \, d\mu$$

$$- \frac{a}{\sqrt{2\pi}} \int_{+1/2\sigma - \frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2} \, d\mu \qquad \text{(A.14)}$$

Differentiating this with respect to Ps ,

$$\frac{dE[Pw;Ps]}{dPs} = \frac{k}{\sqrt{2\pi}} \int_{-1/2\sigma - \frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2} \, d\mu$$

$$+ \frac{kPs}{Ps\sigma\sqrt{2\pi}} e^{-1/2\left[-1/2\sigma - \frac{\ln Ps+\ln k-\ln a}{\sigma}\right]^2}$$

$$- \frac{a}{Ps\sigma\sqrt{2\pi}} e^{-1/2\left[+1/2\sigma - \frac{\ln Ps+\ln k-\ln a}{\sigma}\right]^2}$$

$$= \frac{k}{\sqrt{2\pi}} \int_{-1/2\sigma - \frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2} \, d\mu \qquad \text{(A.15)}$$

The slope of the expected value curve and the slope of the asymptote are equal when $Ps = \infty$. And,

$$\left.\frac{dE[Pw;Ps]}{dPs}\right|_{\infty} \quad \frac{k}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{-1/2\mu^2} \, d\mu = \frac{k}{\sqrt{2\pi}}$$

Therefore, the slope of the asymptote is k .
Also

$$\frac{d^2E[Pw;Ps]}{dPs^2} = \frac{k}{Ps\sigma\sqrt{2\pi}} e^{-1/2\left[-1/2\sigma - \frac{\ln Ps+\ln k-\ln a}{\sigma}\right]^2} > 0 \text{ for all } Ps > 0.$$

Therefore, the slope of the expected value curve is as indicated in Figure 4. By the method of rotating axes, the intercept of the asymptote with the Ps axis can be proved to be $\frac{a}{k}$.[1]

When we considered the price of leverage (P_e) and the price the investor is willing to pay for the warrant (V) , we found,

1. See Kenneth O. May, Elementary Analysis, (Wiley, New York, 1952), pp. 378-80.

$$V = (1-P_e)E[Pw;Ps] + \frac{dE[Pw;Ps]}{dPs} PsP_e \quad . \tag{A.16}$$

Substituting equations A.14 and A.15, the explicit expression for V is,

$$V = (1-P_e)\left[\frac{kPs}{\sqrt{2\pi}} \int_{-1/2\sigma-\frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2}\, d\mu \right.$$

$$\left. - \frac{a}{\sqrt{2\pi}} \int_{+1/2-\frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2}\, d\mu \right]$$

$$+ \frac{kPsP_e}{\sqrt{2\pi}} \int_{-1/2-\frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2}\, d\mu$$

$$+ \frac{kPs}{\sqrt{2\pi}} \int_{-1/2-\frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2}\, d\mu$$

$$- \frac{(1-P_e)a}{\sqrt{2\pi}} \int_{+1/2-\frac{\ln Ps+\ln k-\ln a}{\sigma}}^{\infty} e^{-1/2\mu^2}\, d\mu \tag{A.17}$$

The slope of the asymptote is again k and the intercept of it with the Ps axis is $\frac{a(1-P_e)}{k}$.

APPENDIX B

The maximum likelihood method was used to obtain the estimates of σ , k , and P_e . The actual observed price in the market (Pw) will equal the price the marginal investor is willing to pay for the warrant (V) plus an error term (ϵ) which is assumed to be distributed normally about V with a variance of σ_ϵ^2 . If σ_ϵ^2 were large and Ps were small this might be a poor assumption since V + ϵ must always be greater than zero, but the assumption of normally distributed ϵ implies that V + ϵ can be negative sometimes. By assuming ϵ to be lognormally distributed this problem is avoided but at the cost of added complexity. For the data used σ_ϵ^2 is always quite small and Ps relatively large so the bias introduced by the assumption of normally distributed ϵ will be small.

With this assumption the method of maximum likelihood is equivalent to least squares. The likelihood function, $\Phi(\sigma, k, P_e, \sigma_\epsilon)$ is:

$$\Phi(\sigma, k, P_e, \sigma_\epsilon) = \Phi = \prod_{n=1}^{N} \frac{1}{\sigma_\epsilon\sqrt{2\pi}} e^{-1/2\left[\frac{Pw-V}{\sigma_\epsilon}\right]^2} \tag{B.1}$$

and $\ln\Phi = \Phi^* = -n \ln \sigma_\epsilon - \frac{n}{2} \ln 2\pi - 1/2\left[\frac{Pw-V}{\sigma_\epsilon}\right]^2$

Let $\beta_1 = \sigma$, $\beta_2 = \underline{\overline{P_e}}$, $\beta_3 = k$

Then the normal equations are of the form:

$$\frac{\partial \phi^*}{\partial \beta i} = \frac{1}{\sigma_\epsilon^2} \Sigma \left[(Pw-V) \frac{\partial V}{\partial \beta i} \right] \qquad (B.2)$$

and

$$\frac{\partial^2 \phi^*}{\partial \beta_1 \beta_j} = \frac{1}{\sigma_\epsilon^2} \Sigma \left[(Pw-V) \frac{\partial^2 V}{\partial \beta_1 \partial \beta_j} - \left(\frac{\partial V}{\partial \beta_1} \right) \left(\frac{\partial V}{\partial \beta_j} \right) \right] \qquad (B.3)$$

All empirical work was done under the assumption that the investor thinks stock prices are lognormally distributed. In this case V is given by equation (13). If we make the following notational simplifications,

$$Q_1 = \frac{1}{\sqrt{2\pi}} \int_{-1/2 - \frac{\ln Ps + \ln k - \ln a}{\sigma}}^{\infty} e^{-1/2\mu^2} \, d\mu \ ,$$

$$Q_2 = \frac{1}{\sqrt{2\pi}} \int_{+1/2 - \frac{\ln Ps + \ln k - \ln a}{\sigma}}^{\infty} e^{-1/2\mu^2} \, d\mu \ ,$$

and

$$Z = \frac{1}{\sqrt{2\pi}} e^{-1/2 \left[1/2\sigma - \frac{\ln Ps + \ln k - \ln a}{\sigma} \right]^2}$$

$$\alpha = \ln Ps + \ln k - \ln a$$

Then the explicit expressions for the $\frac{\partial V}{\partial \beta_i}$ are as follows:

$$\frac{\partial V}{\partial P_e} = aQ_2 \ , \quad \frac{\partial V}{\partial \sigma} = aZ \left[1 - \frac{P_e}{\sigma} \left(1/2\sigma + \frac{\alpha}{\sigma} \right) \right] \ , \quad \frac{\partial V}{\partial k} = PsQ_1 + \frac{aP_eZ}{k\sigma}$$

and the $\frac{\partial V}{\partial \beta_1 \partial \beta_j}$

$$\frac{\partial^2 V}{\partial P_e^2} = 0 \qquad \frac{\partial^2 V}{\partial \sigma^2} = aZ \left(\frac{2P_e\alpha}{\sigma^3} - \left[1/4 \frac{\alpha^2}{\sigma^3} \right] \left[1 - \frac{P_e}{\sigma} \left(1/2\sigma + \frac{\alpha}{\sigma} \right) \right] \right)$$

$$\frac{\partial^2 V}{\partial k^2} = \frac{aZ}{k^2\sigma} \left[1 - \frac{P_e}{\sigma} \left(1/2\sigma + \frac{\alpha}{\sigma} \right) \right] \quad \frac{\partial^2 V}{\partial P_e} - \frac{aZ}{\sigma} \left(1/2\sigma + \frac{\alpha}{\sigma} \right)$$

$$\frac{\partial^2 V}{\partial P_e k} = \frac{aZ}{k\sigma} \qquad \frac{\partial^2 V}{\partial \sigma \partial k} = \frac{aZ}{k\sigma} \left(\left[1 - \frac{P_e}{\sigma} \left(1/2\sigma + \frac{\alpha}{\sigma} \right) \right] \left(1/2\sigma - \frac{\alpha}{\sigma} \right) - \frac{P_e}{\sigma} \right)$$

These expressions can be substituted in equations (B.2) and (B.3) to get explicit equations for the $\frac{\partial \phi^*}{\partial \beta_1}$ and $\frac{\partial^2 \phi^*}{\partial \beta_1 \partial \beta_j}$. These equations are non-linear in the parameters and thus it is necessary to use some form of iterative procedure. The method chosen was the method of scoring or Newton's method.[1] Let $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ be the initial values of β_1, β_2 and β_3 chosen in some suitable manner. The method used here was to assume that $\beta_1 = P_e = 0$ and $\beta_3 = k = 1$ since it was felt that their true values were close to these assumed initial values. For each observation from a sample (or alternatively from a randomly chosen sub-sample of observations from the original sample), Pw was assumed exactly equal to V. A σ_n was thus obtained for each observation and the mean σ was chosen as the initial estimate of $\sigma(\hat{\beta}_2)$. Since obtaining the σ_n involves an iterative process, it would seem best to use only a sub-sample of the observations and to get estimates of σ_n good only to two or three significant digits.

Having obtained initial estimates the method of scoring proceeds as follows: Let $\hat{\beta}_1 + \Delta\hat{\beta}_1, \hat{\beta}_2 + \Delta\hat{\beta}_2, \hat{\beta}_3 + \Delta\hat{\beta}_3$ be the true values. Then,

$$0 = \sum_{n=1}^{N} \left((Pw - v) \frac{\partial V}{\partial \beta_1} \Big|_{\beta_1 + \Delta\beta_1, \ \ldots, \ \beta_3 + \Delta\beta_3} \right) \tag{B.4}$$

$$\therefore 0 = \sum_{n=1}^{N} \left(\left[Pw - V(\hat{\beta}_1 + \Delta\hat{\beta}_1, \ \ldots, \ \hat{\beta}_3 + \Delta\hat{\beta}_3) \right] \frac{\partial V(\hat{\beta}_1 + \Delta\hat{\beta}_1, \ \ldots, \ \hat{\beta}_3 + \Delta\hat{\beta}_3)}{\partial \beta_1} \right) \tag{B.5}$$

But

$$V(\hat{\beta}_1 + \Delta\hat{\beta}_1, \ldots, \hat{\beta}_3 + \Delta\hat{\beta}_3) = V(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) + \sum_{i=1}^{3} \Delta\beta_i \frac{\partial V}{\partial \beta_i \partial \beta_j} \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \tag{B.6}$$

and

$$\frac{\partial V(\hat{\beta}_1 + \Delta\hat{\beta}_1, \ldots, \hat{\beta}_3 + \Delta\hat{\beta}_3)}{\partial \beta_1} = \frac{\partial V}{\partial \beta_i} \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} + \sum_{i=1}^{3} \Delta\beta_i \left[\frac{\partial V}{\partial \beta_i \partial \beta_j} \right] \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \tag{B.7}$$

Substituting (6) and (7) in equation (5) we arrive at,

$$0 = \sum_{n=1}^{N} \left[Pw - V(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) - \sum_{j=1}^{3} \Delta\beta_j \left(\frac{\partial V}{\partial \beta_i} \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \right) \right] \left[\frac{\partial V}{\partial \beta_1} \Big|_{\beta_1, \beta_2, \beta_3} + \sum_{j=1}^{3} \Delta\beta_j \left(\frac{\partial V}{\partial \beta_1 \partial \beta_j} \right) \Big|_{\hat{\beta}_1, \hat{\beta}_2} \right]$$

$$= \sum_{n=1}^{N} \left(\left[Pw - V(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) \right] \left[\frac{\partial V(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)}{\partial \beta_1} + \sum_{j=1}^{3} \Delta\beta_j \left(\frac{\partial V}{\partial \beta_1 \partial \beta_j} \right) \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \right] \right. \tag{B.8}$$

1. C. R. Rae, Advanced Statistical Methods in Biometric Research, (New York, John Wiley and Sons, 1952) pp. 165-172.

$$- \sum_{j=1}^{3} \left[\Delta\beta_j \left(\frac{\partial V}{\partial \beta_j} \right) \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \right] \left[\Delta\beta_i \left(\frac{\partial V}{\partial \beta_1 \partial \beta_j} \right) \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} + \frac{\partial V}{\partial \beta_i} \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \right] \Bigg)$$

$$= \sum_{i=1}^{3} \frac{\partial \Phi^*}{\partial \beta_j} \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} + \sum_{j=1}^{3} \Delta\beta_j \frac{\partial^2 \Phi^*}{\partial \beta_i \partial \beta_j} \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3}$$

$$- \sum_{j=1}^{3} (\Delta\beta_j)^2 \left(\frac{\partial V}{\partial \beta_j} \right) \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \left(\frac{\partial V}{\partial \beta_1 \partial \beta_j} \right) \Big|_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \qquad \text{(B.9)}$$

It is assumed that $(\Delta\beta_j)^2$ is negligible and thus the last term in (B.9) drops off. In matrix form let Φ_i^* be the column vector of first derivatives, Φ_{ij}^* be the matrix of second derivatives both evaluated at $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ and $\Delta\beta_i$ be the column vector of changes to be made away from the initial estimates. Then,

$$\Phi_{ij}^* \, \Delta\beta_i = - \Phi_i^*$$

$$\Delta\beta_i = - \Phi_{ij}^{*-1} \, \Phi_i^*$$

The negative inverse of the matrix of second derivatives $\left(- \Phi_{ij}^{*-1} \right)$ gives large sample estimates of the variances and covariances of the $\hat{\beta}_i$. The $\Delta\beta_i$ are added to the $\hat{\beta}_i$ and the whole process repeated until they are small in relation to their variances. The process was stopped when all the $\frac{\Delta\beta_i}{\beta_i} \leq .01$.

As explained in Section 12 the results obtained using the above method were unsatisfactory in that there was usually either a failure to converge or a convergence on what would seem to be on a priori grounds inadmissible values of the β_i. Therefore, a variant of the above method was tried. The initial estimates of the β_i used were the same, i.e., $\hat{k} = 1$, $\hat{P}_e = 0$, $\hat{\sigma} = \sum_{n=1}^{N} \frac{\sigma}{N}$. Instead of iterating to obtain new estimates of k, P_e, and σ, k was assumed to equal one--its initial estimate. The iterative process then involves only two parameters. The derivatives and cross derivatives are easily altered to take into account this assumption. This method also proved to be unsatisfactory because of lack of convergence. The final method involved iterating for only one parameter. In all cases k was assumed to equal to one, but for each sample P_e was assumed equal to any of several different values. For each assumed value of P_e, final values of σ were obtained. The iteration for σ alone involves only the first and second derivatives of the likelihood function with respect to σ, and the variance of becomes $- \frac{1}{\frac{\partial^2 \Phi^*}{\partial \sigma^2}}$

Summary: This article is an extract from a longer presentation submitted as a dissertation to the Department of Economics of Yale University. Its purpose is to investigate the extent and effect of monetary debasement in France from 1295 to 1395. An historical investigation, omitted from this extract, of the institutions and political circumstances of France and England was undertaken in order to establish the economic interrelationships of the period so that theoretical economic concepts with historical validity might be evolved for use in the analysis of monetary debasement. With some reservations the quantity theory of money seemed most useful and it was employed within the framework of an international comparison between France and England. Geographical areas were chosen in the two countries so as to assure similar fluctuations in agricultural output. Price statistics for wheat in the two countries were assembled and the prices were reduced to monetary units of constant silver value. The resultant adjusted price series were found to fluctuate closely enough to allow the conclusion that monetary mutation was swiftly reflected in prices. Since the French price series after adjustment for extensive debasement paralleled the English series adjusted for minor debasement, it was concluded that French prices rose in direct proportion to debasement.

This essay was submitted to Yale University in 1960 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics at Yale.

Dissertation Committee: Professors Robert S. Lopez (Chairman), William J. Fellner, John E. Sawyer.

*

I am indebted to Professors William Fellner, John E. Sawyer and particularly to Professor R. S. Lopez for their invaluable comments. My thanks are also due to the Carnegie Corporation of New York for providing the funds necessary to complete this study.

H. A. M.

PRICE MOVEMENTS AND SPECIE DEBASEMENT IN FRANCE, 1295-1395

Harry A. Miskimin

1. Introduction

The dissertation from which this essay is extracted attempted to investigate the monetary history of France from 1295-1395. Although the nature of fourteenth century documents and material renders illusory the theoretical elegance which might be brought to a contemporary problem of similar nature, the quantity theory of money as expressed in Irving Fisher's exchange equation provides a useful frame of reference. Since the data of the period prohibit quantification of either a velocity or an output term, a second country, England, was introduced in order, through the device of comparison, to control the influence of fluctuations in the output of primarily agricultural societies.

Stripped of historical context and qualifications, an economic rationale or model, of a purely hypothetical nature, justifying this approach could be developed as follows. The economies of two countries, say France and England, are defined in terms of the Fisher tautology, $MV = PT$, where T is the real output of the single product, wheat. This produces two equations.

England $\quad M_e V_e = P_e T_e$

France $\quad M_f V_f = P_f T_f$

Assume that there is only one determinant, climate, of the T terms and that the countries are so situated that climate is the same in each country.[1] It follows that, simultaneously with any change in T_f, T_e would move in precisely the same proportion and direction, i.e., that $T_e/T_f = K_1$, a constant. Assume further that the V terms also behave in a similar manner so that $V_e/V_f = K_2$, a constant.

Given these assumptions regarding the V and T terms, we may form the following equation:

$$\frac{M_e}{M_f} \cdot K_2 = \frac{P_e}{P_f} \cdot K_1$$

It is then possible to measure the effect of a change in M_f upon P_f by comparing the latter to P_e if it be given that M_e remains constant. Since the V and T ratios are constants the entire impact of a change in M_f falls upon P_f. The comparison of P_f and P_e, however, comprehends changes in the T terms since any change in T_f is assumed, because of climate to be paralleled by a change in T_e.

Both economies are assumed to operate on a silver standard without fiat money.[2] The governments establish the account value of silver by law in minting coins. As a result of this, the M terms must be divided into two parts: A, the number of units of account to the mark weight of silver; and B, the total number of marks of silver in the country. If the B terms are constants or behave similarly, i.e., if $B_e/B_f = K_3$, a constant,[3] the

---

1. The historical validity of this assumption is discussed in Section 3.

2. The historical validity of this assumption is discussed in Section 2.

3. The degree to which the quantities of bullion in the two countries actually

equation then becomes:

$$\frac{A_e}{A_f} \cdot K_3 \cdot K_2 = \frac{P_e}{P_f} \cdot K_1$$

or

$$\frac{P_f}{F_f} \cdot \frac{K_3 \cdot K_2}{K_1} = \frac{P_e}{A_e}$$

and substituting

$$K_4 = \frac{K_3 K_2}{K_1}$$

$$\frac{P_f}{A_f} \cdot K_4 = \frac{P_e}{A_e}$$

If this latter equality is true no matter what changes occur in A_f, it follows, since P_e/A_e is independent of A_f, that French prices, P_f, rose in direct proportion to debasement, i.e., to the change in A_f.

Further if $P_f/A_f = P_e/A_e$ then $K_4 = 1$.

Since P_f, A_f, P_e and A_e are measurable within the limits imposed by the data of the fourteenth century, it is possible to test, in effect, the proposition that $K_4 = 1$. Graph 4-1 shows the result of such a test and tends to support this proposition,[1] and hence, the correlative statement that French prices rose in direct proportion to monetary debasement.

Now $$K_4 = \frac{B_e}{B_f} \cdot \frac{V_e}{V_f} \cdot \frac{T_f}{T_e}$$

and as already mentioned there is historical evidence for the proposition that climate did tend to make $T_e/T_f = K_1$, a constant,[2] and, although of a weaker order, historical evidence that B_e/B_f may broadly have behaved as a constant. If this evidence were strong, the conclusion that $V_e/V_f = K_2$, a constant, would be justified but, due to the nature of the sources, it would seem more wise to restrict the scope of the conclusion to the more modest statement that possibly $K_4 = 1$ or at the most that there is some evidence

---

meet these assumptions was investigated in Chapter V of the dissertation. Working only with the amount of metal coined each year, it would appear that over the course of the century the bullion supplies of the two countries fluctuated similarly. In the short run, there is some evidence, although of a weak order, that there may have been some "normal" requirement for bullion in each country. These requirements, in ratio form, may have behaved as a constant in the short run although enormous fluctuations of the actual quantity certainly occurred around the so-called "normal" requirement.

1. The reader is again reminded of the numerous qualifications imposed by the events of history and the nature of the data.

2. See Section 3.

that $(B_e/B_f)(V_e/V_f)$ is a constant.

As the reader is no doubt aware, the model described above is initially a device for overcoming the absence of data adequate for the study of one country alone. No pretense is made that such a model may be applied in a purely functional manner to the history of the fourteenth century.

In the dissertation itself, a general discussion of the historical ethos, including an investigation of the legislation, political history, institutions, production methods and population was conducted in order to evaluate the nature of the economic relationships of the period and specifically to inquire into credit and the velocity of circulation. The latter remained to a substantial degree one of the unknowns although some circumstantial historical evidence tends to support the proposition that velocity behaved similarly in the two countries.

The present excerpt assumes that the reader has accepted the historical arguments concerning credit and velocity and also the rationale for using only grain as the commodity whose price is to be measured. Problems involving the B terms, i.e., the quantity of bullion, are omitted from this essay but are considered in the dissertation. I hope that the reader who wishes to question this essay on general historical grounds or for its treatment of credit, velocity and the quantity of bullion will consult the arguments to be found in the dissertation.

2. The Monetary Systems of England and France

The unit of account in fourteenth century France was the livre, divided into twenty sols, composed of twelve deniers. This monetary system, descended from the time of Charlemagne when the Libra was in reality a pound,[1] continued in France until the Revolution. It still exists in England in the system of pounds, shillings and pence. In both countries, however, the units merely indicated a relationship within the sub-multiples of the system and, by the fourteenth century, had ceased to correspond to the metalic pound weight.

In both countries, the king determined the metallic value of the unit of account by specifying the number of livres or pence which must be struck from the respective metallic units of weight. In France, the standard used was the mark of Paris or Troyes of 4,608 grains. In England, the measure was the Tower Pound of 5,400 grains.[2] During earlier periods of French history, there were at least four different marks but, in the reign of Philippe I, the mark of Paris was selected and its use was extended to the entire kingdom.[3] Perhaps because of the earlier diversity of weights, two units of account exist simultaneously in France. These are the livre Parisis and the livre Tournois. Although both are founded upon the mark of Paris during our period, the livre Parisis is one fourth heavier in metal content than the livre Tournois. This, since the mark if the same, is only a question of face value. I have converted all money and prices, when specific reference is made to the unit of payment, into the Tournois system. Any future reference to livres, sols or deniers should be understood to apply to the Tournois system unless the alternate system is specified.

Deferring for the moment the question of the alloy as commanded by

---

1. A. Dieudonne, Monnaies royales Francaises depuis Hugues Capet jusqu'à la revolution, Man. de Numismat. Franc. vol. II, Paris: 1916, p. 79.

2. John Craig, The Mint, Cambridge University Press, Cambridge, 1953, p. xv.

3. Dieudonne, op. cit., p. 39.

the king, it is necessary to attempt to resolve the more elusive question of the alloy of the so-called fine metal to which the legally specified alloy was added.[1] Coin fineness in England was designated as 926 parts in one thousand. In France coinage silver, called argent-le-roy, was to be eleven and one half parts fine out of twelve or 958.18 in one thousand. Dieudonne refers to some analyses of the Gros Tournois, a common French coin, which show argent-le-roy to be actually only 941 grains fine.[2] The difference between the standards of the two countries, taken at legal value, would be 32 parts in one thousand or 3.2 per cent. If English silver matched the penny trial plate of 1279, it was in reality 923.3 fine.[3] The difference between the two coinage bases would then be 18.7 parts in one thousand or 1.87 per cent if computed on the basis of the actual fineness. Craig estimates that the touchstone, even in the hands of an expert, was almost certainly incapable of distinguishing differences in fineness more closely than ten parts in one thousand.[4] We may therefore assume that the difference between the silver bases of the two countries would have been barely distinguishable by a contemporary observer. For the purposes of this study, I have not attempted to adjust for this slight difference between the two countries' standards of fineness but have accepted them as equal. This produces a small error but one which is probably not greater than the alternative of accepting either the legal weight or of trying to devise an estimate of the varying deviation from the legal weight. It offers the advantage of maintaining the integrity of the series for each country and averts the possibility that too great an insistence on accuracy might convert the figures for each country into inaccurate abstractions. Since, in the market, the difference would be scarcely observable, the economic error involved would be negligible.

Similar treatment is accorded to the standards for gold. Subject to the technical limits of refining skill, the gold mark and the gold pound were both fine gold. The variation between the two countries depends upon techniques of refining but since gold was the standard for international transactions, the degree of variation is probably smaller than that inherent to the silver base of the two countries.

In France, the silver coinage was measured and controlled by the pied de monnaie, a number which expressed the relationship between the titre or alloy, the face value, and the weight of the coin. Following de Saulcy,[5] if n is the number of coins to the mark of silver, d the face value of the coin, and t the titre, expressed in deniers, the pied is the following relation:

$$\frac{n \cdot d}{5 \cdot \frac{t}{12} \cdot 12} = \text{Pied de monnaie}$$

The denominator may appear a bit mysterious. Silver is considered fine if it contains 12/12 of silver, argent-le-roy, i.e., silver 23/24 fine,

1. Craig, op. cit., p. 52.

2. Dieudonne, op. cit., p. 39.

3. Craig, op. cit., p. 52.

4. Ibid., p. 15.

5. F. de Saulcy, Recueil de documents relatifs à l'histoire des monnaies frappées par les Rois de France, 4 vols., Paris, 1879, vol. 1, p. xiv.

so that t/12 is the manner of expressing the amount of alloy added to argent-le-roy. The division by 12 is required to convert the face value, d , from deniers into sols. There is to my knowledge no explanation of the division by five.

The pied de monnaie is, in itself, an index of debasement since it accounts for all possible variations in the currency except for those of design or nomenclature which have no economic significance. For our purposes, however, it will, perhaps seem more familiar to speak of the number of livres drawn from the mark. This latter figure is simply one fourth of the pied de monnaie and naturally varies in the same manner as the pied, reflecting all changes of titre, face value or weight. The pied, and hence the number of livres to the mark, will always refer to the mark of silver, argent-le-roy, because of the division by the titre and not to the marc d'oeuvre or mark of coin used for internal transactions was normally silver with gold reserved for large or international transactions of either a political or a trading nature.[1] The French system was not bimetallism in the usual sense of the word but was, rather, a silver standard supplemented by a valuable commodity, the gold coin. It was not possible to take silver to the mint in order to demand payment in gold. The coins were not commercially exchangeable for each other at par.

The king, in issuing the silver coinage, designated the number of livres to the mark since he established the face value of the coins concerned. This was also attempted for the gold coinage but it frequently proved infeasible. An independent market rate existed for the gold coins[2] and this fluctuated with the pied de monnaie of the current silver coinage. Since the market rate for gold or gold coins was rendered in livres, sols or deniers with a silver base, we may, I think, safely consider the gold coinage as a commodity and relate all problems of price to the silver coinage alone. The rationale for this approach is fortified by the fact that the king himself, in periods of debasement, paid for gold in the gold money coined and without reference to its value in money of account.[3] Since the mints paid for the bullion in the coin which they struck from it, there could be no possibility for an arbitrage situation to develop between the two metals.

In France there are two important sources for information regarding the coinage. One is the collected ordinances by which the king communicated his will regarding the pied de monnaie to the masters of the money. In some cases, this will was specific in defining all features of the coin but in others it simply expressed the pied and allowed the masters to determine the specific form of the actual coin.[4] De Wailly, in 1857,

---

1. Marc Bloch, Esquisse d'une histoire monétaire de l'Europe (Cahier des Annales, N° 9) Paris: A. Colin, 1954, p. 39.

2. N. de Wailly, "Mémoire sur les variations de la livres tournois depuis le regne de Saint Louis jusqu'à l'établissement de la monnaie decimale," Mem. de l'Inst. Imp. de France, Acad. des Inscript. et Belles Lettres, vol. 21, 1857, pp. 408-19 (1343-61).

3. See for example, Lettres Patentes, 15 June 1359 in de Saulcy, op. cit., p. 406; see also A. Landry, Essai économique sur les mutations des monnaies dans l'ancienne France de Philippe le Bel à Charles VII, Paris: H. Champion, 1910, p. 156.

4. Landry, op. cit., p. 16.

working primarily with these ordinances, prepared a list of the variations in the metal content of the livre for both the gold and the silver coinages.[1] This list is subject to some flaws of omission in the fourteenth century and also in later periods.[2] In addition to this defect, it tends to convey a false conception of the effective dates of the pieds involved since the actual coins usually follow after and sometimes even precede the dates of the ordinances.

The alternative to this source is the collection of the mint records which has been compiled by de Saulcy.[3] While this work includes ordinances and relatively modern commentaries upon the early French coinage, it also contains most of the documents which remain to us regarding the activities of the fourteenth century mints. These are unfortunately incomplete since fires in 1737 and 1776 and thereafter the Revolution destroyed a large part of the documents of the Chambre des Comptes. The documents seem most complete in the periods, 1308-12, 1338-39, and after 1350, but in most other years some information is available. Where they are markedly incomplete I have accepted the figures from the ordinances and compared the result with the tables printed in de Wailly.

The most common mint documents printed in de Saulcy are those which arise from the requirement that the coins struck be assayed by the masters of the money in Paris. The individual mintmaster, under the supervision of a royal guard, were required to send a sample of the coinage, selected at random, to the general masters in Paris where the assay was made before the coins remaining could be put into circulation. The size of the sample was, for gold, usually one coin for each five hundred struck and, for silver, one coin for each thousand. The documents typically give the name of the mint, the period referred to and a statement of the face value, the titre and the number of coins to the mark. They then give either the number put in the box for assay, the number struck or occasionally the number of marcs d'oeuvre. Often both the number struck and the number in the box was given.

I have used these mint documents in preference to the ordinances for two reasons. First, since they provide the dates of the actual coinage, we may, in effect, see the simultaneous existence of several pieds or numbers of livres to the mark. We are, therefore, not led into the erroneous belief that the pied from the ordinances, which is unique to its date, is also exemplified in the actual coinage. For a study of the reaction of prices to debasement, this becomes quite important. The second reason for preferring this group of documents is, of course, that they yield more information. They are the only source from which it is possible to estimate the output of the mints. For the specific problem of debasement, we may, I think, be quite confident in accepting as complete the series here presented since it comprehends both ordinances and mint documents.

I have utilized the table printed in Craig's work, The Mint, for the construction of the English debasement series and for the output of the London mint. The English data are more complete than the French and have been sufficiently studied to be accepted, at least for our purposes, without much hesitation. No unresolved questions remain regarding the

---

1. de Wailly, op. cit., passim.

2. See Marc Bloch, "Prix, monnaies, courbes," Ann. Ec., Soc. civ. N° 4, 1946, pp. 355-57.

3. de Saulcy, op. cit.

infrequent and well-documented changes in the metal content of the pound.

From the formula for the pied de monnaie, it is apparent that there are three variables through which a mutation of the coinage may take place. The face value can be raised or lowered, the weight can be altered by taking more or fewer coins from the mark, and the alloy can be raised or lowered. From an economic point of view it is insignificant which one or combination of the three is actually employed. Only the outcome has economic importance and this outcome is measured by the pied or, dividing this number by four, by the number of livres to the mark.

The frequency of mutation of the coinage is significant and, ideally, one would desire to have enough prices to assure the coincidence of price and pied in order to discover the effect of each monetary change upon prices. Unfortunately, as will be apparent in Chapter III, this is not the case. In order to overcome the resultant difficulty, we shall be concerned only with the maximum number of livres drawn from the mark in any year. The lesser numbers of livres to the mark will be considered only for their influence, sometimes substantial, upon the accuracy of our approach to the available prices. In Graph 2-1, I have plotted the maximum number of livres drawn from the silver mark in each year. The additional mutations are plotted simply as points in order to indicate both the number of changes and the range of variation in the coinage in a given year.[1]

Graph 2-2 shows both the number of pounds of account drawn from the silver pound tower and the number of pounds of account drawn from the gold pound Tower. Since there are so few changes in the English money, it has not been necessary to follow the procedure employed in the French series for indicating more than one change in a given year. Even the most cursory inspection of this graph will, I think, prove adequate justification for referring to the English coinage as essentially sound in comparison with the French case. The debasements which do occur increase the number of pounds of account drawn from the pound Tower by a little under twenty-five per cent over the course of the century. One money is normally maintained for long periods of time in contrast to the French practice of issuing as many as twenty varieties, all with different metallic content, in a single year. The English system does sometimes contain money of different metal contents but this is usually to cover the additional cost of the fractional coinage and the difference between the two intrinsic values is slight. From 1335 to 1343, 252 and 254 pence were taken from the Tower pound. From 1344 to 1351, the range was 266 to 270 pence. These variations are not important enough to be considered but, in accordance with the method employed for the French series, I have always worked with the maximum number of pence to the pound Tower.

The gold coinage poses particularly complex difficulties in the case of France. I have argued that gold should be treated as a commodity and not as money, since it was valued in livres and sols of an account system founded upon a silver standard and since the mints normally returned gold coins for gold bullion brought to the mint. The mints often list the gold coins which they strike at the face value which they were declared to have at the time of the last preceding ordinance, despite the fact that the market rate existing at the moment of coinage may have been much higher. To use

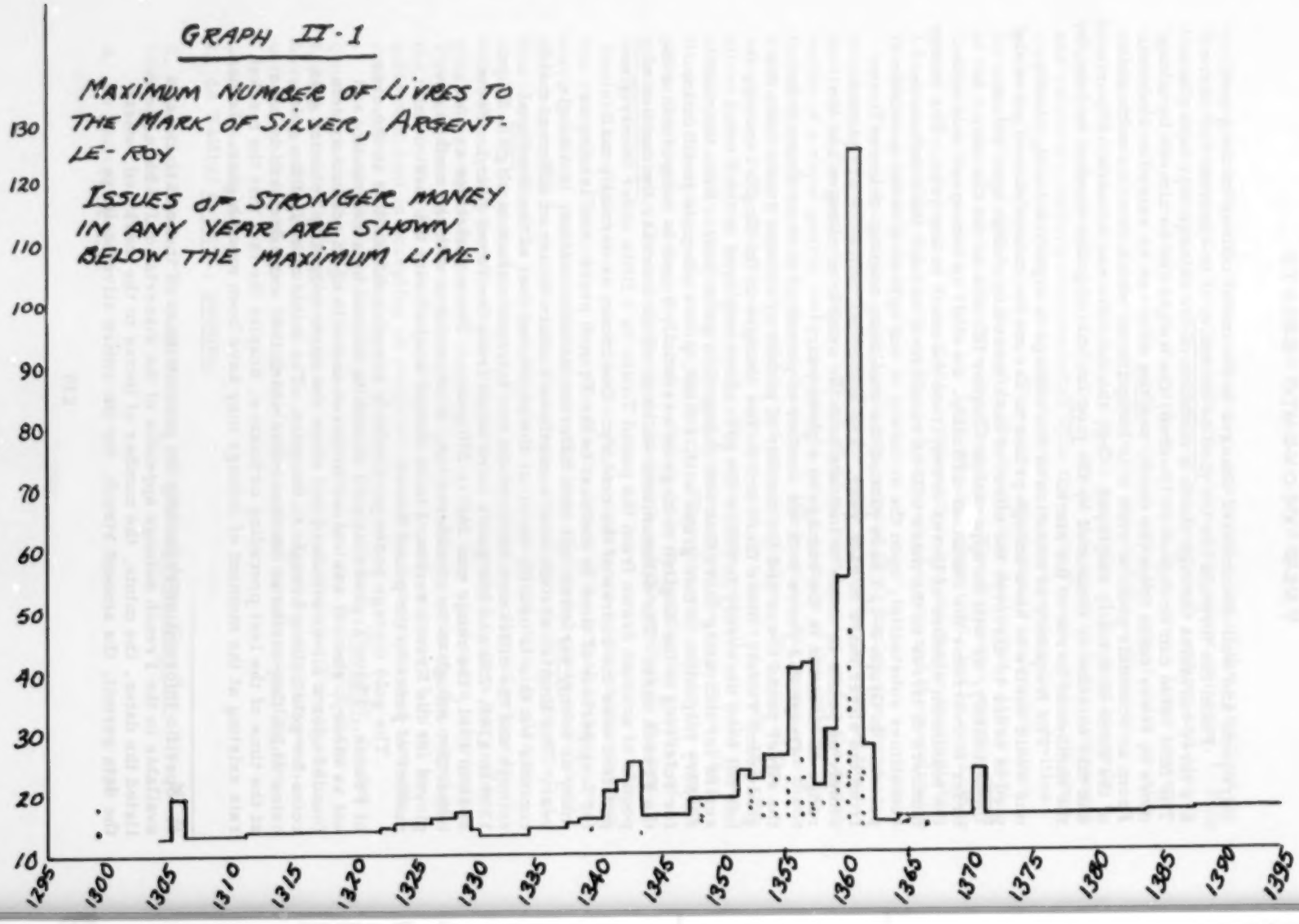
---

1. Specific information regarding the precise dates of these variations is available in the French coinage appendix of the dissertation. I have there listed the dates, the mints, the number of livres to the mark and, where the data permit, the amount struck, for the entire silver coinage.

GRAPH II-1

MAXIMUM NUMBER OF LIVRES TO THE MARK OF SILVER, ARGENT-LE-ROY

ISSUES OF STRONGER MONEY IN ANY YEAR ARE SHOWN BELOW THE MAXIMUM LINE.



GRAPH II-2

GRAPH II-2

NUMBER OF POUNDS OF ACCOUNT TO THE ENGLISH POUND TOWER.

UPPER SCALE REFERS TO GOLD COINAGE

LOWER SCALE REFERS TO SILVER COINAGE

this official figure to establish a gold-silver ratio would clearly be an anomaly since no exchange of gold for silver took place at this rate. On the other hand, the use of the market rate for gold coins does not provide us with an official figure for the gold coinage, and suffers from the additional defect of being subject to the vagaries of time, place and market conditions which the normal short-run conception of the mint gold-silver ratio does not include. The latter factors might force a change in an official gold-silver ratio, but they should not be used to compute one.

For the sake of obtaining an official figure for the number of livres to the gold mark, I have used the most recent official face value for each gold coin as it is issued and have plotted the result as Graph II-3. The reader is again warned, however, that this is to some extent an arbitrary procedure and that it must not be used to calculate either the effective or the officially enforced gold-silver ratio. Official gold debasements will appear as increments in the graph but this does not necessarily indicate the purchasing power of the gold coin in terms equivalent to those of the silver-based money of account. The gold bullion does exist as an alternative to silver as a means of payment for some transactions but it would be valued in silver standard units of account. Estimates of the quantity of gold issued will be provided in Section 5 and will be expressed in marks of fine gold.

In England in 1351, it was declared "lawful for every man to exchange gold for silver, or silver for gold, or for gold and silver; so that no man hold a common exchange, nor take profit for making such exchanges, upon pain of forfeiture of the money so exchanged, except the king's exchangers, who were allowed the profit of such exchanges, according to the ordinance aforetime made."[1] It was also declared that plate should be received at the mints by weight and that the money, when made, should be delivered by weight.[2] The first ordinance would seem to indicate that a strong attempt was made in England to maintain the legal tender quality of gold and silver coins or, in other words, to maintain the official gold-silver ratio which was, in effect, abandoned in France during periods of extreme debasement. The payment be weight at the mints in conjunction with the profit on the exchanges sets limits to the exchange of gold for silver or vice-versa. It also indicates, if there is a profit to be had, that the official gold-silver ratio is not an equilibrium one. This would seem consistent with Feaveryear's conclusion that gold was not domestically an alternative to silver as a means of payment in England.[3]

The various prohibitions against the export of money, with the exception of the gold noble,[4] would limit the possibilities of international arbitrage in the money metals. Although the official English gold-silver ratio may, perhaps, be considered partially effective domestically, it was not, by itself, a factor which would have been capable of initiating silver flows. Gold could leave the country but, if foreign silver was brought in, it had to be taken to the mint for recoinage. The profit of the kings' exchanges would presumably stop the conversion back into gold. I do not mean to suggest by

GRAPH II-3

1. Rogers Ruding, Annals of the Coinage of Great Britain and Its Dependencies, Third ed., London, 1840, Vol. 1, p. 227.

2. Idem.

3. A. E. Feaveryear, The Pound Sterling. A History of English Money, Oxford: Clarendon Press, 1931, p. 22.

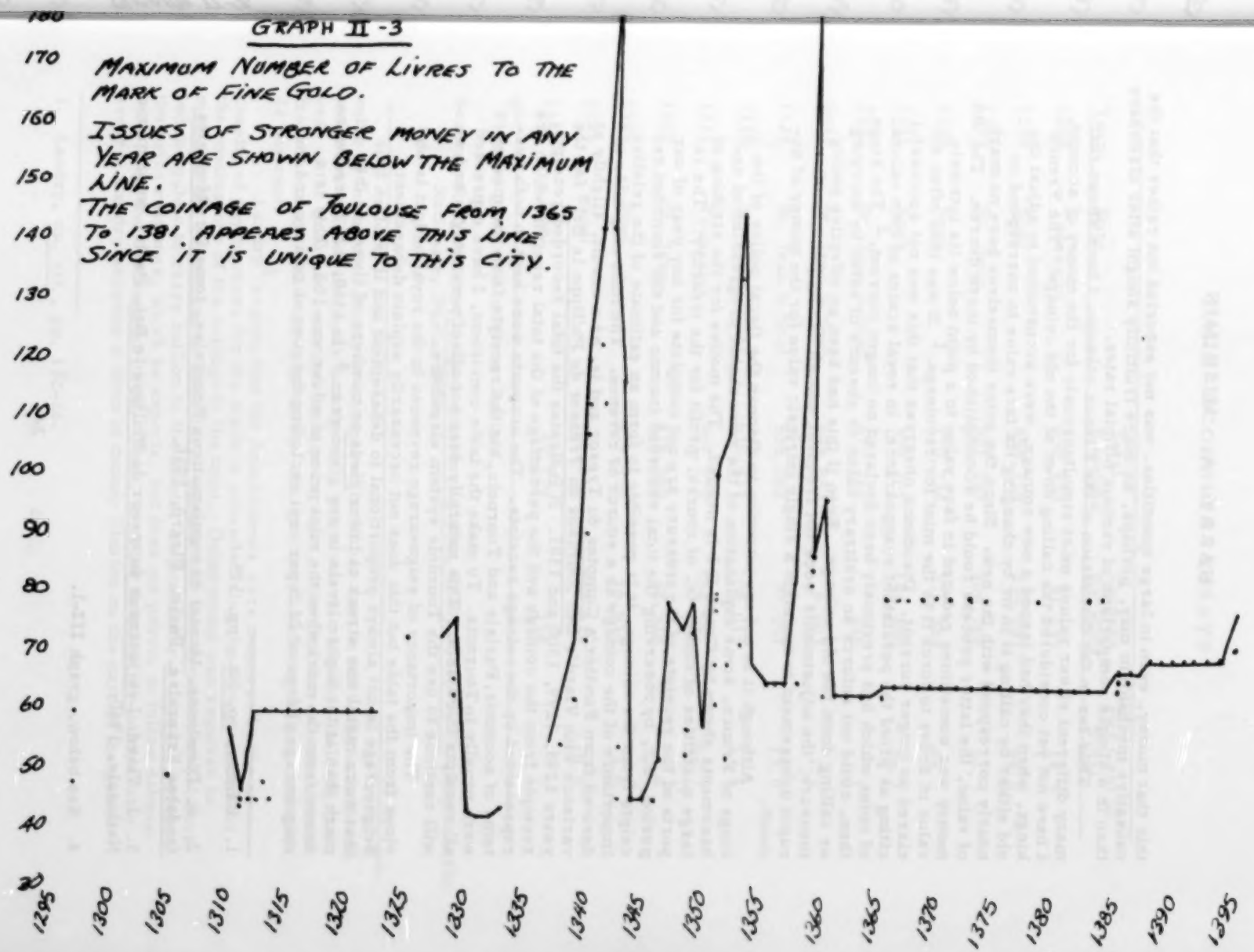
4. Ruding, Annals of the Coinage, Vol. 1, pp. 215, 216, 221.

GRAPH II-3

MAXIMUM NUMBER OF LIVRES TO THE MARK OF FINE GOLD.

ISSUES OF STRONGER MONEY IN ANY YEAR ARE SHOWN BELOW THE MAXIMUM LINE.

THE COINAGE OF TOULOUSE FROM 1365 TO 1381 APPEARS ABOVE THIS LINE SINCE IT IS UNIQUE TO THIS CITY.



this official figure to establish a gold-silver ratio would clearly be an anomaly since no exchange of gold for silver took place at this rate. On the other hand, the use of the market rate for gold coins does not provide us with an official figure for the gold coinage, and suffers from the additional defect of being subject to the vagaries of time, place and market conditions which the normal short-run conception of the mint gold-silver ratio does not include. The latter factors might force a change in an official gold-silver ratio, but they should not be used to compute one.

For the sake of obtaining an official figure for the number of livres to the gold mark, I have used the most recent official face value for each gold coin as it is issued and have plotted the result as Graph II-3. The reader is again warned, however, that this is to some extent an arbitrary procedure and that it must not be used to calculate either the effective or the officially enforced gold-silver ratio. Official gold debasements will appear as increments in the graph but this does not necessarily indicate the purchasing power of the gold coin in terms equivalent to those of the silver-based money of account. The gold bullion does exist as an alternative to silver as a means of payment for some transactions but it would be valued in silver standard units of account. Estimates of the quantity of gold issued will be provided in Section 5 and will be expressed in marks of fine gold.

In England in 1351, it was declared "lawful for every man to exchange gold for silver, or silver for gold, or for gold and silver; so that no man hold a common exchange, nor take profit for making such exchanges, upon pain of forfeiture of the money so exchanged, except the king's exchangers, who were allowed the profit of such exchanges, according to the ordinance aforetime made."[1] It was also declared that plate should be received at the mints by weight and that the money, when made, should be delivered by weight.[2] The first ordinance would seem to indicate that a strong attempt was made in England to maintain the legal tender quality of gold and silver coins or, in other words, to maintain the official gold-silver ratio which was, in effect, abandoned in France during periods of extreme debasement. The payment be weight at the mints in conjunction with the profit on the exchanges sets limits to the exchange of gold for silver or vice-versa. It also indicates, if there is a profit to be had, that the official gold-silver ratio is not an equilibrium one. This would seem consistent with Feaveryear's conclusion that gold was not domestically an alternative to silver as a means of payment in England.[3]

The various prohibitions against the export of money, with the exception of the gold noble,[4] would limit the possibilities of international arbitrage in the money metals. Although the official English gold-silver ratio may, perhaps, be considered partially effective domestically, it was not, by itself, a factor which would have been capable of initiating silver flows. Gold could leave the country but, if foreign silver was brought in, it had to be taken to the mint for recoinage. The profit of the kings' exchanges would presumably stop the conversion back into gold. I do not mean to suggest by

---

1. Rogers Ruding, Annals of the Coinage of Great Britain and Its Dependencies, Third ed., London, 1840, Vol. 1, p. 227.

2. Idem.

3. A. E. Feaveryear, The Pound Sterling. A History of English Money, Oxford: Clarendon Press, 1931, p. 22.

4. Ruding, Annals of the Coinage, Vol. 1, pp. 215, 216, 221.

GRAPH II-3

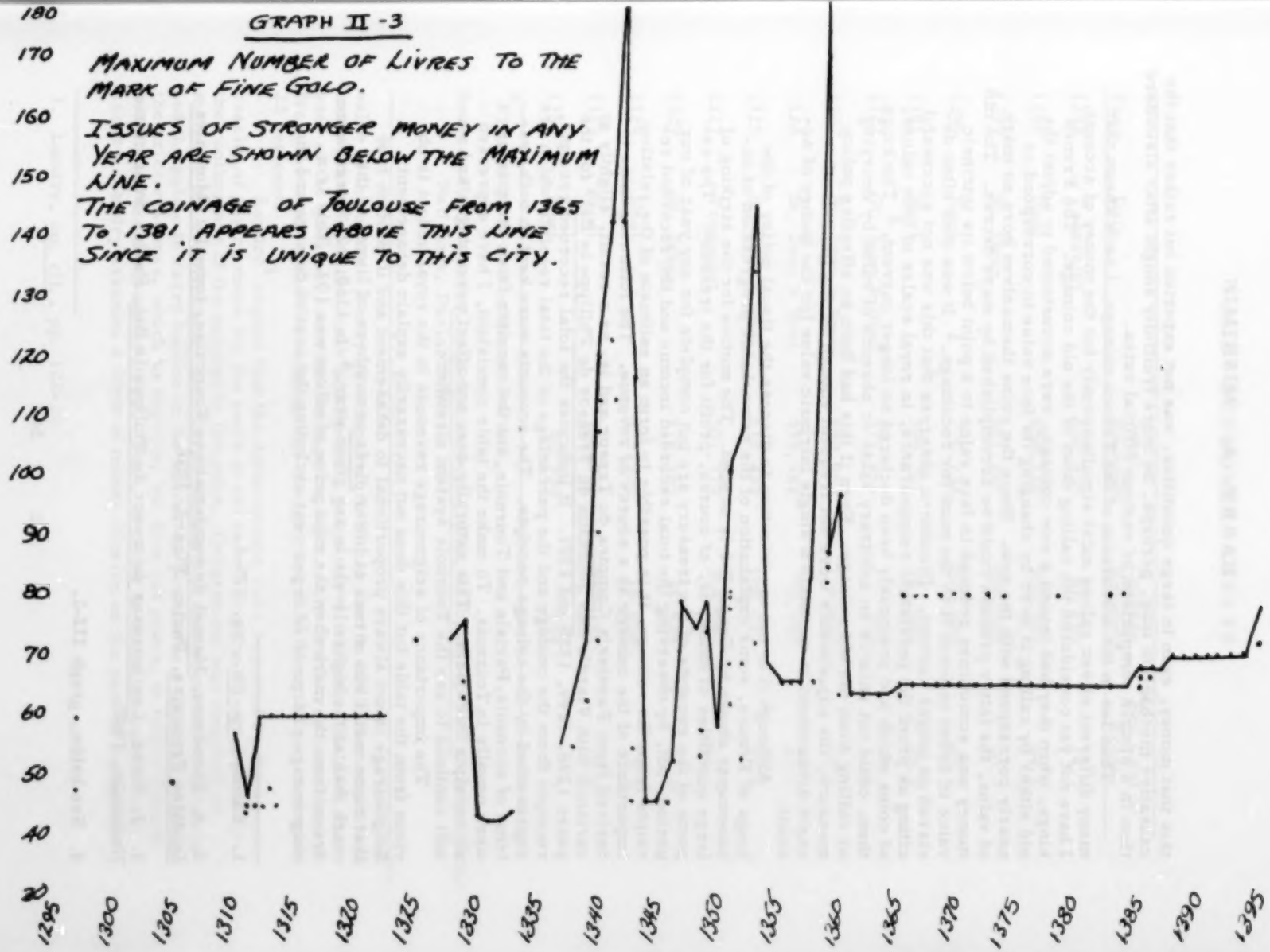
190

GRAPH II -3

MAXIMUM NUMBER OF LIVRES TO THE MARK OF FINE GOLD.

ISSUES OF STRONGER MONEY IN ANY YEAR ARE SHOWN BELOW THE MAXIMUM LINE.

THE COINAGE OF TOULOUSE FROM 1365 TO 1381 APPEARS ABOVE THIS LINE SINCE IT IS UNIQUE TO THIS CITY.



this that money, even in large quantities, was not exported but rather that the causative mechanism may, perhaps, be more fruitfully sought after elsewhere than in a simple comparison of various official rates.

Thus far in the discussion of the French coinage, I have shown that many different silver values exist simultaneously for the money of account. I have not yet considered the calling down of the old coinage. The French kings, when they had issued a new coinage, were accustomed to adjust the old either by calling it in or by changing its face value to correspond or nearly correspond with the new. Since the coins themselves bore no mark of value, the latter process could be accomplished by mere decree. The old money was sometimes reduced in face value to a point below its intrinsic value in order to force it to the mint for recoinage.[1] It was then often declared no longer current. Dieudonne observes that this was not successful citing as proof the persistent reappearance, in royal scales of coin values, of coins which had previously been declared no longer current.[2] The kings, then, could not enforce an arbitrary value or absence of value by decrying or calling down the old coinage. Even if this had been an effective policy measure, the adjustments were not frequent enough during periods of rapid debasement to maintain a single intrinsic value for the money of account.

Although it is not my intention to discuss the fiscal policy of the kings of France, some explanation of the remarkable progression of debasements shown by Graph II-1 is needed. The motive for the striking of large quantities of money is, of course, profit for the treasury. The reports of the receipts of the treasury are not complete for any year of our period, but, by observing the total recorded income and the recorded receipts from the coinage, it is possible to form an estimate of the relative importance of the coinage as a source of revenue. The following table is derived from Fawtier's Comptes du Tresor and is, as a result, slightly at variance with Viard's Les Journaux du Treasor de Philippe le Bel[3] for the years 1298, 1299, 1300 and 1301. It indicates the total recorded receipts, receipts from the coinage and the percentage of the total receipts which are represented by the coinage receipts. The accounts were kept in both systems of accounts, Parisis and Tournois, but the receipts from the monies were usually in Tournois. To make the table consistent, I have expressed all receipts in Parisis. This naturally does not affect percentages and we will continue to use the Tournois system elsewhere.

The importance of seigneurage revenues in the royal budget is obvious from the table but this does not necessarily explain debasement. Seigneurage is not always proportional to debasement and it is often true that more metal was struck at lower pieds or numbers of livres to the mark than at the higher levels in any given year.[4] In 1360, 125 livres were drawn from the markwhen the mint price of silver was 192. This is a seigneurage charge of 22.5 per cent excluding the cost of coinage and the

1. Landry, op. cit., pp. 175-6.

2. A. Dieudonne, Manuel de numismatique Francaise, tome IV: Monnaies feodales Francaise, Paris: Picard, 1936.

3. J. Viard, Les journaux du tresor de Philippe le Bel, Paris: Imprimerie Nationale, 1940.

4. See below, graph III-1.

Year	Date	Total Receipts *	Coinage Receipts *	Per Cent of Total
1298	Toussaint	550,923	81,149	14.7
1298	Noel	858,319	555,097	64.6
1299	St. Hean	801,243	406,708	50.7
1299	Noel	770,734	376,284	49.2
1301	Noel	399,201	144,490	36.7
1322	St. Jean & Noel	477,366	546	.1
1323	St. Jean & Noel	598,052	6,603	1.1
1324	St. Jean & Noel	538,382	70,453	13.0
1325	St. Jean & Noel	610,437	108,013	17.6
1327	Toussaint	501,779	314,691	62.7
1329	St. Jean	379,585	28,528	7.4
1329	Noel	459,950	16,636	3.6
1349	Noel	781,746	552,028	70.6
1384	Noel	47,779	334	.7

*Figures given are in livres Parisis.

Source: R. Fawtier, Comptes du tresor (1296, 1316, 1384, 1477). Imprimerie Nationale, Paris, 1930.

value of the alloy. It may be compared to 30 per cent in 1338 when six livres were drawn from the mark and the mint price was 4.6 livres. In 1349, the year of the greatest recorded coinage receipts, only nine livres were drawn from the mark while the seigneurage charge ranged between 50 per cent and 22 per cent.

Landry[1] argues that the debasements were required by the necessity of paying more for the mark in order to bring the silver needed for the continuance of the coinage to the mint. Debasement was required to maintain the seigneurage. In order to explain why this sequence would have led the public to carry bullion to the mints, he postulates a time lag between the price response, which he expects, but does not prove, to follow debasement, and the debasement itself. Then, in order to make his point doubly secure, Landry assumes a form of money illusion on the part of the public

1. Landry, op. cit., pp. 122-26.

which leads them to expect that prices paid in strong money will be the same as those transacted in weak money.[1]

My argument against this thesis is the subject matter of Sections 3 and 4 where it is shown that the accounting systems of the period and the behavior of prices seem inconsistent with any public illusions regarding the metallic content of the coinage. Graph III-1 in the following section shows that less metal was coined into debased money than into sound money even during periods of extensive debasement. This would seem to indicate that the public was neither conscious of any substantial benefit to be obtained through lagged price rises nor afflicted with any large degree of money illusion.

As an alternative to Landry's thesis, I would suggest that a rise in the mint price of the silver mark did bring some additional money or metal to the mint but for different reasons than those which he offers. On the one hand, a rise in the mint price without debasement, i.e., a reduction in seigneurage, increases the real mint price and might bring more silver to the mints. On the other hand, a number of payments, feudal contracts, rents, etc., were fixed in terms of the money of account and could be paid in either debased or sound coin. Debasement might have caused people to bring sound money to the mints in order to receive bad money for the payment of such obligations. If, for example, a good denier contained enough silver to make six bad ones and the seigneurage charge was one third, a man bringing the good denier to the mint would have received four of the debased coins in return. He might then pay a contractual obligation of one denier and have three deniers left after the transaction. The three remaining deniers, even if they purchased only one half as much as the one good denier, would be clear profit. The king would have had a profit of two bad deniers and the landlord would have been taxed five-sixths of the real value of his accustomed revenue.

Marc Bloch suggests that the king could augment the number of units of payment in his hands by coining his own silver into debased money and using it to pay his obligations.[2] This has several drawbacks. In taking this action, the king would reduce the real value of his own fixed revenue. He did not normally possess a great stock of metal. His contracts were often fixed in terms of gold or foreign money as in the case of loans and, finally, prices rose against him. This motive for debasement may, as a result of these qualifications, be considered minor.

The mechanism which I have offered as a possible explanation for a flow of metal to the mint in response to debasement is only tentative. It seems as reasonable as others suggested in the literature and, in the absence of any concrete evidence, is as empirically justified. It is consistent with the violent protests of both the landlords and the clergy and through them provides a further possible motive for debasement. In the early fourteenth century, the French kings had almost no permanent taxes and were compelled to raise money for the tremendous expense of almost constant war by means of lay and clerical contributions. Debasement could be used as a means of extracting additional payments from those to whom contractual payments were important. The choice open to this class was no longer between granting or not granting a royal subsidy but rather, because of the possibility of debasement, between being ruined

1. Idem, p. 124.

2. Bloch, Esquisse, p. 66.

by the decrease in the real value of their revenue or of paying enough taxes to secure sound money. On 11 May, 1305, for example, Pope Benedict XI gives Philippe le Bel a significant portion of the church revenues in order to aid him in the return to strong money.[1] In February, 1357, the estates are willing to raise a subsidy which they would administer in return for being allowed to regulate the coinage.[2]

The increase in expenses entailed by the war explains, I think, the necessity of employing such drastic measures to raise revenue. The nature of military operations may also account for the emphasis placed upon monetary manipulation since this was the only method, other than loans which were also used, for mobilizing large sums quickly. The king could, by means of debasement, rapidly obtain an increase in both bullion and units of account and, at the same time, force certain segments of the population to grant decimes and subsidies in order to protect their fixed revenues.

If my explanation of the public motive for bringing silver to the mints is correct, there is a limit to the effectiveness of continuous debasement as a measure for raising seigneurage profits. Payments specified in money of account were of limited magnitude and tended to decline in number as means of avoiding the loss of purchasing power were sought. As the intrinsic value of the money of account diminished, less silver would have been required to pay all of the existing contractual obligations. If, for example, the sum of such payments was one thousand deniers, one mark of silver would be required for the payments when one thousand deniers were drawn from the mark. If a debasement caused ten thousand deniers to be drawn from the mark, only one tenth of a mark of silver would be required to meet the assumed contractual payments of one thousand deniers. If this debt were the only motive causing the public to bring silver to the mints, less and less metal would have been brought as debasement increased and less metal would have been coined at peaks of debasement. This is consistent with the evidence regarding the output of the mints.

In the early part of the fourteenth century, the decime was, perhaps, the most important source of revenue, after seigneurage. While, in form, it still had to be granted by the clergy, it had, in fact, become almost regular. There was a decime in every year from 1285 to 1328 and at least twenty between the latter year and 1350.[3] The remainder of the revenues consisted of lay subsidies, centimes, cinquintimes, vingtimes, a temporary hearth tax in 1304, etc.[4] In 1341, the royal salt monopoly was extended to all the towns of France thus providing an additional source of revenue.[5] Such revenues, although often substantial, were slow of col-

1. de Saulcy, Recueil, p. 161.

2. A. Coville, Histoire generale, histoire du Moyen Age, tome VI, l'Europe Occidentale de 1270 a 1380, Duexieme Partie 1326-1380, Presses Universitaires de France, 1940, p. 555.

3. G. Dupont-Ferrier, Etudes sur les institutions financieres de la France a la fin du Moyen Age, Paris, Firmin-Didot et Cie, 1930, 1932, Vol. I, p. 135.

4. R. Fawtier, Histoire generale, histoire du Moyen Age, tome VI: l'Europe Occidentale de 1270 a 1380, Premiere partie de 1270 a 1328, Paris, Presses Universitaires de France, 1940, pp. 202-3.

5. Coville, Hist. Gen., pp. 524, 603.

lection and, with the exception of the salt tax, required a certain amount of time to arrange before the process of collection could even be begun. Since war prevented the kings of France from accepting this loss of time, the coinage was of prime importance and had to be manipulated for revenue purposes until a more satisfactory financial system could be devised.

The change took place after the middle of the century. The capture of the French king, Jean le Bon, by the English at Poitiers caused fiscal reform in France in order to meet the ransom of three million gold ecus. This reform generated an administration which became permanent.[1] By 1369, Charles V benefited from both sides and fourages which were permanent until the end of the century,[2] if one excepts the brief but severe tax struggle following Charles' death in 1380 and continuing until 1384.[3] It is, I think, not a coincidence that the period of sound money is cotemporal with the new and permanent taxes.

Since the coinage was a major source of revenue during much of the century, it will be worth while to consider some of the bounds upon its effectiveness. Adequate treasury reports would be of inestimable value but, unfortunately, the only years for which such reports exist are years where the coinage records are lost or show only a small volume. This obviously prevents us from making a direct comparison of the profit from the coinage and the output of coins. It is not possible to compute the coinage revenue from the mint documents since the mints list neither the sources nor the prices of their raw materials. If the writings of Nicholas Oresme, monetary theorist and adviser to Charles V, be taken as proof, the mints, at their discretion, might sometimes have purchased one metal with the other.[4] This would further compound the difficulty of making an estimate of the revenue from a given volume of coinage, since it precludes any simple calculation based upon the difference between the mint price of the mark and the number of livres drawn from it.

Even in the absence of direct evidence, we may identify two important restrictions on the extent of seigneurage profit. These were the inability of the kings to force bullion to the mints and the problems arising from the competition of other monies. The former has already been mentioned. Declarations that specific coins were no longer current were designed to cause bullion to flow to the mints but these have been shown to have been of limited effect. It has also been suggested that raising the price of bullion at the mints would and, in fact, did bring only diminishing quantities of silver to the mints for coinage. Obvious impediments, both political and quantitative, would have blocked the continuous, effective use of such ordinances as that of 1337 requiring that half the bessels of gold and silver in the kingdom be carried to the mint.[5]

---

1. Idem, p. 584.

2. Idem, p. 601.

3. Joseph Calmette and Eugene Deprez, Histoire generale, histoire du Moyen Age, tome VII, la France et l'Angleterre en conflit, Paris: Univ. de France, 1937, pp. 11-18.

4. Nicholas Oresme, De Moneta, ed. Ch. Johnson, London: Thos. Nelson, 1956, p. 15.

5. E. Perroy, "Les crises du XIV^{e} Siecle," Annales economies, societies, civilizations, 1949, No. 2, p. 176.

The highly developed international trade which continued throughout the century also imposed limits upon the profit which could be drawn from the coinage. Trade was carried on in such commodities as wool, cloth, silk, furs, jewelry, robes, armor, spices, wine, salt, fish and others. Trade routes connected the Mediterrean countries with each other and with the countries of Central Europe and the Baltic.

It is not my present intention to summarize this trade but it is important to recognize it as a possible determinant of the degree to which the coinage could be exploited for revenue purposes. The familiar injunctions against the domestic use of foreign money were at once, attempts to defend the royal prerogative of coinage against foreign competition and to force the foreign specie to the mint for recoinage. The presence of foreign money and the existence of substantial trade tended to make the public sensitive to mutations of the royal money since it was frequently evaluated against the foreign money in the course of trade.[1] This, naturally, would have restricted the king from purchasing goods with debased money at the prices which existed under sound money. International trade also presented the problem of an adverse balance of payments. Mint profit would decline if bullion left the country and was not available for coinage.

In preparation for the discussion of grain prices, it is necessary to outline the trade in this commodity. In the medieval period, the greatest international commerce in grain was centered in the Mediterrean where Sicilian and Spanish grain in conjunction with that transported by Genoa could be cheaply carried by water. Southern France was at times an importer of grain. In 1311, 1333, and 1344, we have records that show the importation of wheat by merchants of Montpellier, Nimes and Narbonne.[2] In the latter year it reached the rather impressive figure of fourteen thousand setiers.[3] In 1329, wheat was imported for the use of the inhabitants of Carcassone; in 1337, the export of wheat from England to Gascony was licensed.[4] Both the French and the English-dominated portions of southern France seem to have alternated between a precarious balance and a net deficit in their domestic grain production.

England and northern France, which later supplied large quantities of grain to the Netherlands,[5] were surplus areas and exporters. An exception to this pattern is found in the years of the great and almost universal famine of 1315-17 when Edward II sought to import grain from northern France and finally obtained it from the above-mentioned southern European sources, Genoa, Sicily and Spain.[6]

Commerce in grain did not normally exist between England and France,

---

1. Bloch, Esquisse, p. 66.

2. M. Larenaudie, "Les famines en Languedoc aux XIV^{e} et XV^{e} Siecles," Annales du Midi, 1952, pp. 27-39.

3. Idem, loc. cit.

4. Idem, p. 29.

5. M. M. Postan, "Some Social Consequences of the Hundred Years War," Economic History Review, Vol. XII, 1942, pp. 1-12.

6. H. S. Lucas, "The Great European Famine of 1315, 1316 and 1317," Speculum, Vol. V, October 1930, pp. 343-377.

defining the latter country in political and not geographical terms. Edward III prohibited the export of corn to any place but Ireland and required licenses for export to the English possessions of Calais and Gascony.[1] Direct trade between the two political units was so impeded by the war[2] that it may, I think, be neglected. This is particularly true for grain where little comparative advantage was evident between the two countries and where political restrictions abounded. Other trade, especially that carried on through Italian intermediaries, cannot be denied but for the purposes of our price survey, grain prices, in those portions of each country with which we shall be concerned, are primarily domestic prices, indigenous to the individual country. The export trade in grain is a small portion of the domestic trade, subject to the political restrictions of supplying the home market first. The two countries export, in most cases, to widely different geographical areas although a partial exception is found in the proximity of Calais, an English market, and the Netherlands, a French market. The possibility of interconnection of English and French prices through common export markets, particularly when given the hazards and costs of medieval trade, is, I think, remote enough to be ignored.

International trade has been shown to have been a factor which increased the public's sensitivity to monetary change and which represented some degree of competition to the royal monopoly of the coinage. Feudal coinages had also been a threat and an impediment to the royal monetary prerogative and the long and continuous effort which the French kings, since the beginning of the thirteenth century, had devoted to the elimination of such money may serve as an indication of the degree of importance which they attributed to their own coinage rights. The history of the acts and regulations regarding the feudal coinages has been summarized in so convenient a form by Duby in his Traite des monnaies des barons[3] that I shall take the liberty of providing an English precis of the relevant passages.

> Louis VII forced the seigneurs who wished to alter their money to advise the king four months in advance so that his subjects might dispose of the old money and thus suffer no loss. This is confirmed by a charter of the Bishop of Meaux dated in May, 1225. P. xxxiv.
>
> In 1226, Louis VII declared his money to be current over the entire realm while feudal money was limited to its place of origin. P. xxxiv.
>
> Saint Louis in 1262 prohibited any baron who did not enjoy coinage rights from using any but the king's money and confirmed the ordinance of 1226. P. xxxiv.
>
> In 1273, Philippe le Hardi confirmed the ordinance of 1262 and forbade the weakening of royal money. In 1275, he reiterated the prohibition on the coinage of money similar to his own. In 1278, he purchased the coinage rights of the Bishop of Albi. P. xxxv.

---

1. N. S. B. Gras, The Evolution of the English Corn Market from the Twelfth to the Eighteenth Century, Cambridge, Harvard University Press, 1915, p. 135.

2. Alice Beardwood, Alien Merchants in England, 1355-57, Cambridge, Mass.: Med. Acad. of Amer., 1931, n. p. 3.

3. P. -A. Duby, Traite des monnaies des barons, Paris: Imp. Royal, 1790, 2 vols., Vol. 1, pp. xxxiv-xxxix.

In 1289, Philippe le Bel renewed the ordinances of 1226, 1262, and 1273. P. xxxv.

There is a note, circa 1300, that no baron other than the king may make money of either gold or silver or worth more than a denier. P. xxxv.

In 1313, Philippe le Bel prohibited the barons from striking money until they had obtained letters patent from him indicating the titre and manner of manufacture. P. xxxvi.

In June, 1313, he renewed the ordinance of 1289. A royal guard was placed in the baronial mints to assure the delivery of deniers at the ordained price. The guard had to be present at the purchase of gold and silver and support the ordinances of the king. The masters of the royal money must assay the feudal money and enforce the prescribed weight and titre. A baron issuing defective money must remedy the fault and surrender all profit on the issue. P. xxxvi.

On Nov. 28, 1315, the barons agreed to conform to the titre, quantity, and weight set by the Generaux des Monnaies. They also agreed that their money would be local and the royal money national and that their money was subject to royal assay. The Generaux des Monnaies set the total feudal coinage at one thousand marks a year. Nine hundred marks were to be coined into deniers and one hundred into mailles. P. xxxvii-iii.

On May 14, 1319, Philippe le Long purchased the monies of Chartres and Anjou; on January 27, 1320 he acquired those of Clermont and Bourbonois. P. xxxviii.

On May 3, 1328, Philippe le Valois bought the money of Blois. P. xxxix.

It would seem that, by the early part of the fourteenth century, the feudal coinages had been severely limited both in number and in importance. By the beginning of our period, such feudal money as remained was regulated by royal authority and followed the pied established for the royal money. Since the seigneurial coinage was limited to one thousand a year, it could not divert important quantities of bullion from the royal mints and, hence, could not limit their profits. Regulation was enforced by the presence of royal guards and by royal assay so that, despite the fact that thirty feudal coinage rights were extant in 1315,[1] their effect was small both economically and as a hinderance to the king. The influence of feudal coinages upon prices may be considered to have been insignificant by the fourteenth century.

3. Wheat Prices in England and France

The most severe impediment to successful medieval price history is not, perhaps, the absence of data, although this is significant, but rather the omni-present difficulty of interpreting the data which exist. Extreme care must be taken in order to expose those disparities which may be inherent in a medieval statistical series, and, hence, to indicate the range of variation which may be expected to arise from the nature of the data.

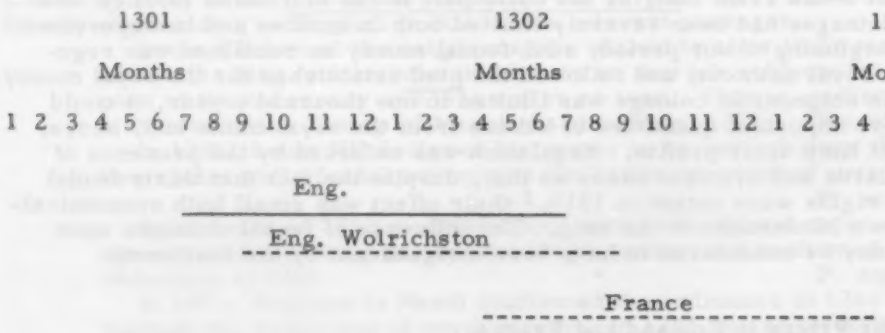
One source of error, particularly serious for a comparative study, is the recurrent problem of establishing precise dates for those transactions undertaken at the prices which remain to us. Records frequently span a defined period, a month, a year, simply listing prices and quantities sold during the interval without reference to the specific sales date.

1. de Saulcy, Recueil, pp. 192-3.

In part this may be the result of the small number of transactions in the period, but it is, I think, more likely that documents pertinent to the original sales were more subject to loss than those records maintained by central organizations in charge of a number of manors. These latter accounts may have been preserved, in England at least, for their value in providing title to lands possessed, whereas the documents related to the original sale of commodities, if such documents existed, would have had little value to commend their safekeeping.[1]

English prices, which I have derived from Thorold Rogers' History of Agriculture and Prices, refer to a year which ran from July to July. An exception to this is found in Wolrichston where the year extends from Michaelmas, i.e., September, to Michaelmas. Since seven months of a given year, say 1301-2, are actually in 1302, prices which are not otherwise identifiable are placed in the latter year. When it has been possible to do so, the precise date has been used to place prices in the proper modern year in order to assure the greatest possible chronological correspondence between the French and English price series.[2]

The difficulties which arose from the lack of dates are apparent. If the harvest of 1302 fails, a maximum price following the harvest would be credited to the following year, 1303, and as a result, be placed one year out of correspondence with French prices despite the possible real congruence of dates. The irregularity of the French year prevents us from simply working in the original dates and further magnifies the possible deviations. This year goes from Easter to Easter or for 1302 in the old system, from April 1302 to April 1303 in the modern calendar. A simple diagram will perhaps be useful in making this variation more clear.



Since prices falling in any of the years indicated by the lines below the months would be placed in 1302 unless otherwise dated, it is clear that the worst possible error would occur if an English price from July, 1301 were compared to a French price from April, 1303, a difference of twenty-two months. Since Easter is a variable holiday, this latter figure is partic-

1. J. E. T. Rogers, A History of Agriculture and Prices in England, 7 vols., Oxford, The Clarendon Press, 1866, Vol. 2, p. xvi.

2. In the dissertation in an appendix on prices, I have included dates when they are available so as to indicate the degree of certainty which may be attributed to this process of dating.

ular to the year 1302 but in all years some error may be expected from this source.

It is fortunate for our purposes that medieval agricultural price variations were very pronounced. Despite the fact that prices are in some instances improperly dated, this does not seem to alter greatly the general shape of the price curve, but rather has the effect of moving the peaks to the left or right by a year. When, in the absence of any other explanation, a French price rise culminates one year before or after an English rise, it is not, I think, unreasonable, particularly where the rise is in both cases steep, to ascribe the non-coincidence to a false date and to accept both increases as the result of identical variations in climate. The only alternative is the unreasonable assumption that there was a one year lag between weather conditions in England and France. Since, for reasons which will follow, we shall be concerned with maximum prices rather than any sort of average, the important movements in the wheat price will not be lost even if they chance to fall in the wrong year.

Seasonal variations will thus be minimized since the highest prices will not be averaged into lower prices of another part of the year. Where prices are few and are not equally distributed over the year, such an average would give undue weight to the season best represented in the data. The maximum price at least has the advantage of being a simple number and is free from these difficulties of interpretation. While this reduces the influence of seasonal disparities, it cannot entirely eliminate them. In years where only one or very few prices are available, there is no way of determining how representative of the actual maximum the highest recorded price might be. In the case, for instance, that a French price occurring after the harvest were compared with an English price before the harvest, significant seasonal fluctuations would bias the comparison.

Regional divergencies in price exist in both countries, but since Rogers' collection is extensive enough to permit the selection of only those prices from the area immediately surrounding Oxfordshire, the regional influence may be virtually eliminated from the English series. N. S. B. Gras, working with Rogers' prices, has divided England into fifteen price areas and has shown that significant movements of prices tend always to be in the same direction throughout all the areas.[1] Despite his questionable statistical practice of taking an average over a period of almost two hundred and fifty years and problems connected with the comparability of the measures used, Gras does, I think, establish the similarity of movement between his districts for large changes. If this is accepted, we are justified in using one area, Oxfordshire and environs, as indicative of the trend throughout England. By choosing the best documented region we at once assure the adequacy of material and eliminate the geographical variations.

In France we are less fortunate. The only substantial collection of prices is that of D'Avenel,[2] a work so thoroughly permeated with inaccuracies as to render it useless save as an index to other sources of prices which have a wide geographical dispersion.

The location of the towns from which prices have been taken may be seen on the accompanying map. The possibility of error from regional varia-

1. Gras, Evolution of the English Corn Market, p. 47.

2. Geo. D'Avenel, Histoire economique de la propriete, des salaires, des denrees, et tous le prix en general depuis l'an 1200 jusqua'en l'an 1800, Paris, 1894, 7 vols.

CALAIS
TOURNAI
ST. QUENTIN
ROUEN
PARIS
BRITTANY
TROYES
CHAUMONT
ANGERS
BOURGES
MONSTREUIL-BONIN
POITIER
SAIMIERS
LA ROCHELLE
ST. POURCAIN
MACON
CREMIEU
LYONS
LIMOGES
DAUPHINE
MONTPELLIER
TOULOUSE
MODERN FRANCE
FRANCE IN 1305
ENGLAND IN 1259
ENGLAND IN 1328
ENGLAND IN 1361
ENGLAND IN 1380
NAVARRE IN 1361
IN 1365-78

Key to Map I

1.	Aisne	Angers
2.	Alluyes	Bourges
3.	Ambleny	Calais
4.	Angers	Chaumont
5.	Archery	Cremieu
6.	Arras	La Rochelle
7.	Artois	Limoges
8.	Avranches	Macon
9.	Beaumont-le-Roger	Montedome
10.	Bellencombe	Montreuil-Bonnin
11.	Caen	Montpellier
12.	Cani	Paris
13.	Chartres	Poitiers
14.	Corbeil	Rouen
15.	Coutances	St. Lo
16.	Dourdan	St. Pourcain
17.	Dragueville	St. Quentin
18.	Eu	Sommieres
19.	Flagny-en-Gatinais	Toulouse
20.	Fontenai-sur-le-Vei	Tournai
21.	Jumieges	Troyes
22.	Genets	
23.	Lattanville	
24.	La Guerche	
25.	Mery-sur-Seine	
26.	Muneville	
27.	Orleans	
28.	Paris	
29.	Provins	
30.	Rouen	
31.	St. Lo	
32.	St. Pair	
33.	St. Sauveur-le Vicomte	
34.	Varenge-a-Oseville	
35.	Vincennes	

---

tions must be acknowledged here, but if they are no wider than those of all England, they may be assumed to be of small magnitude. Beveridge thinks that the use of diverse measures was often responsible for what had formerly been taken to be regional variations in price and suggests that these, in England, may not have been very substantial.[1] The regional error is minimized by the use of maximum prices, since whenever there is a choice, the area of high relative prices will be selected, thus automatically reducing the impact of geographical differences. The use of maximums also tends to magnify price movements and thereby diminishes the relative importance of the comparatively small interregional variations. Price fluctuations due to

1. W. H. Beveridge, "Wheat Measures in the Winchester Rolls," Ec. Hist. Vol. II, 1930-33, p. 44.

climate and output changes are of large enough magnitude, I think, to outweigh the geographical factors.

From the map it is clear that the geographical locations from which prices have been taken are not chosen entirely at random. The price towns lie north of the Loire and are grouped in a fairly wide strip on both sides of the Seine. Prices are not taken from Bretagne, a territory in constant dispute throughout the century, and not in its economic or political structure inherently French. Prices from towns south of the Loire have also been excluded, as the climate of southern France is so unlike the north that the use of southern prices would invalidate the French-English comparison. Beyond this factor, the southern border of France, as may also be seen on the map, is subject to dramatic change as the English possessions expand and contract about Bordeaux, approaching in 1361 almost the third part of France. Despite the fact that these lands were theoretically held in fief from the king of France, the almost constant political and military struggle for them and the assumption by Edward III of the title of King of France argues strongly in favor of omitting the area on political as well as meteorological grounds. In the north, prices from Ponthieu and Calais have also been excluded since these are English possessions during much of the century.

The lands of the king of Navarre are normally included since they are effectively held in fief from the king of France during most of the century. After 1355, when Charles the Bad, king of Navarre, is allied first with Etienne Marcel and then with the English against the Dauphin, these lands are not used as price sources. Paris is not used during the brief period from 1357 to 1359, since, under the leadership of Etienne Marcel, the town is in open revolt against the Dauphin and, while under his siege, usurps the royal coinage rights by striking an unauthorized issue.[1] In 1378, following an attempt to poison him, Charles V seizes all the lands of Charles the Bad with the exception of Cherbourg, which had been given to the English by this seigneur in an attempt to strengthen his position against his enemy, Charles V.[2] After this date, prices have again been taken from these lands since they are free, thereafter, from those political disturbances which might render them distinct from prices from other areas in France.

The quality of the wheat sold at any given price remains, in general, an unknown factor. That qualitative differences existed is beyond doubt but in the absence of evidence to the contrary there is no other course to follow than to assume that all wheat is identical. Where there is evidence that the wheat is of a special type, e.g., seed, the price has been omitted.

In no case have prices been used which might reflect either interest or rent but this also is accomplished only by excluding what is known to be biased and not by the preferable method of including only those prices which are known to be free of this bias. The chance, however, that a wheat price includes interest is probably smaller than that involved in using other commodities, such as wine or wool, which are known to have been the object of forward purchases by the Italians. In England, for example, the Ricardi, the Peruzzi and the Bardi all advanced money on wool of future growth;[3] in France, at Beaucaire, St. -Gilles, Nimes and Lunel,

1. Coville, Hist. Gen., p. 563.

2. Idem, pp. 640-41.

3. M. M. Postan, "Credit in Medieval Trade," Economic History Review, Vol. I, 1928, pp. 234-261.

the Lombards purchased grapes two months before the actual harvest.[1]

Rents of wheat are sometimes sold and D'Avenel has tried to convert these into wheat prices by the simple assumption that they represent capitalized wheat. This process is clearly invalid and only those prices which reflect the sale of commodities are used. One exception must, however, be allowed: the price of wheat was usually free but the price of bread was regulated on the basis of the free market wheat price. For the purpose of establishing such a price the municipal officials would make an estimate of the free price and use this to set the bread price. These are retained in the case of Chartres since they are the most complete series of prices for a single town and are virtually unbroken throughout the century. These prices, unfortunately, are not apt to be the highest prices occurring in any year and, as a result, in many years, fall below the curve of the maximum prices. In the years where these prices are the only ones available, they have been used despite the probability that they introduce a downward bias.

In a sense, the discussion of the price data has so far been conducted in terms applicable to a modern economy where the money price represents the simple exchange of money, say a dollar, for a commodity whose price is a dollar. It is at this point that the comparison fails to hold and that we must investigate the mechanism of the money of account in order to discover what is meant by the statement that the price of a setier of wheat was, e.g., 19 sols, 1 livre, etc.[2]

The previous section has indicated the frequency of monetary mutations. In the year 1360, there were twenty changes in the money and, therefore, twenty different silver contents were established for the monetary units, livre, sol, and denier. The prices which we have are normally expressed simply in terms of these monetary units without reference to the silver content of the livres or sols in which they were paid. Owing to the facts that coins were themselves given a denomination in money of account by royal decree, that prices and records were maintained in a money of account, and that these were not the same although expressed in identical terminology, the monetary price history of the period has sometimes appeared rather confused. It is possible, however, by observing a contemporary account, to make this mechanism more clear and thus, perhaps, to overcome these difficulties of terminology.

The Compte de la Commanderie de Saint-Denis for the period from July 1, 1329 to July 1, 1330 will serve this purpose.[3] In this document we find an account with the following heading:[4]

---

1. E-R. Labande, "L'administration du duc d'Anjou aux prises avec le probleme du blé en Languedoc, 1365-1380," Annales du Midi, 1950, pp. 5-14.

2. Luigi Einaudi, "The Theory of Imaginary Money from Charlemagne to the French Revolution," in Enterprise and Secular Change, ed., F. C. Lane and J. C. Riemersma, Homewood, Illinois: R. D. Irwin, 1953, passim; Dieudonne, Monnaies Royales, pp. 77-81.

3. Printed in Germaine Lebel, Histoire administratif, economique, et financier de l'abbaye de St.-Denis, Paris, 1935.

4. Lebel, Hist. admin. p. 374.

Pro expensa coquinarii media et forti moneta reducta
ad debilem

III^{m} $IIII^{xx}$ II lb XVIIIs IId

The preceptore, one Fratre G. de Fontaneto, by this entry indicates the type of money in which the accounts are to be kept and states that stronger monies have been converted into weak money for the sake of the necessities of bookkeeping. A little further along another entry provides information as to how this is done.[1]

per eundem xxii modii vi sextarii, xi lb iis pro modio medie
monete et xix mod., vii lb xvix pro medio fortis, valent
vi^{c} xxix lb viiis debil.

In this entry the market price is first given in terms of the decreed value of the coin used for payment and then reduced to a common money of account after the sale so that it may be recorded in a common manner. The higher price in "medie monete" and the lower price in "fortis" informs us that the mechanism employed for adjusting to debasement is to increase the price of the commodity rather than to reduce the value in money of account of the coins at the time of the sale. The latter alternative would make prices meaningless as a measure of the reaction to debasement.

With each new issue of silver money, the king established the account value of the coin struck. In adjusting an old coinage or in founding a new one, the usual penalty, that his commands be followed "sur peine de cors et d'avoir,"[2] indicates that the king expected to be obeyed. Since neither party to a sale was injured by paying or being paid in debased money, as long as the price adjustment was complete and provided that there was no delay in payment, there is little reason to believe that the procedures of St. -Denis were not general to France. Nothing was to be gained and much was to be lost by holding the price of a commodity constant and reducing the face value of debased money, if the same thing could be accomplished legally by taking the money at par and varying the price.

In the previous section we have shown that, although the ordinances establishing the pied de monnaie do not overlap in time, the actual coinage frequently does, during periods of extreme debasement or rapid alteration of the coinage. While this fact alone is sufficient evidence for the simultaneous existence of several royal coinages, it may be useful by way of further proof to anticipate our consideration of the output of the mints.

The accompanying chart shows the presently known total of silver marks, argent-le-roy, struck in each year from 1354 to 1361. Below this total, the thin lines indicate the number of marks struck at each pied. The numbers at the extremities of these thin lines represent the number of livres drawn from each mark.

From this graph we may read the number of mutations in any year and their relative importance in terms of metal struck. Since some of the new coinage is the reminting of older coinage, it is important to know to what extent we may expect the older coinages to disappear in this reminting process. If, on the graph, the volume of real silver struck in any year in the most debased coin were greater than that struck in the better monies of lower pieds, one could argue that the early money had been, or might

1. Idem, p. 375.

2. de Saulcy, Recueil, p. 162; Ordonnances, Vol. I, p. 432; de Saulcy, p. 60; Ordonnances, Vol. I, p. 429.

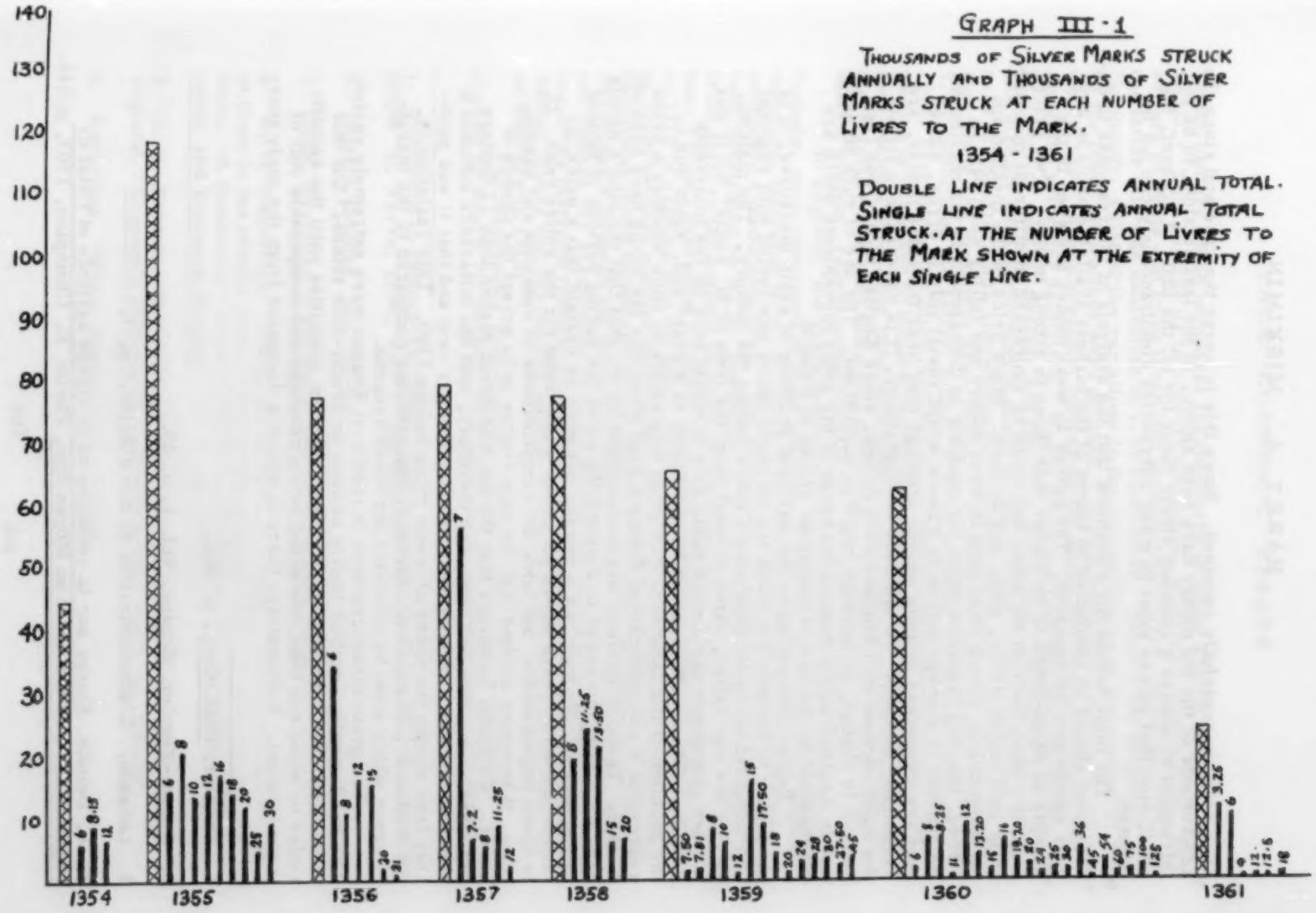
GRAPH III-1
140

GRAPH III · 1

THOUSANDS OF SILVER MARKS STRUCK ANNUALLY AND THOUSANDS OF SILVER MARKS STRUCK AT EACH NUMBER OF LIVRES TO THE MARK.

1354 - 1361

DOUBLE LINE INDICATES ANNUAL TOTAL. SINGLE LINE INDICATES ANNUAL TOTAL STRUCK AT THE NUMBER OF LIVRES TO THE MARK SHOWN AT THE EXTREMITY OF EACH SINGLE LINE.



have been, successfully recoined. Since this is never the case and since the adjustments of the old money have been shown, in the last section, to be too infrequent to assure a constant silver value for all the units of account, the conclusion that prices exist in many currencies simultaneously can not be doubted.

The logic behind my statement that the analysis of the response to debasement should be conducted in terms of the highest available price in any year should now be apparent. The prices in weak money are seen in the Compte de Saint-Denis to be higher than those in strong money. Two prices from the same town on the same day could be expressed in different monetary units both called livres, sols, and deniers. Since price is the mechanism for adjustment, rising when paid in weak money and falling when paid in strong, the only plausible price for analysis of debasement is the maximum. An attempt to average prices in France would result only in error, since they are expressed in units which, although they bear the same names, are essentially distinct. To maintain comparability, it is necessary to follow the same method with England even though, since English money was less subject to change, an average could be significant. The use of such an average in England would dampen the range of the price movements which are emphasized by the use of maximum prices in France.

Before proceeding to the intricate problem of grain measures, it is necessary to mention the sales tax on wheat which was levied with regional and temporal irregularities over France during the century. The maltote, as this tax was called, dates at least from the time of St. Louis[1] but it was neither permanent nor always applicable to wheat even by the fourteenth century. The duration varies from a month to a year[2] and, again due to the problem of the calendar, it is often impossible to know whether a price represents a sale contracted during a time when this tax would have been effective. These difficulties are compounded by the fact that the king's agents had, and used, the power to suspend the sales tax during periods of famine or hardship.[3] We may, as a result, not even be certain that a price includes the tax in years where all goods were taxed for the entire year. If, to these impediments, one adds the complications of evasions and exemptions, it becomes evident that the only course is to accept the chance of error. We know, however, that the tax was never higher than 12 deniers per livre, i.e., five per cent, and therefore, that the difference between a tax free sale and a taxed sale is at most five per cent and that it was probably less during the years of lower rates before 1360. This variation is, for medieval price history, virtually insignificant compared to the margin of error which must be tolerated for other reasons.

The grain measures used in medieval France were extremely varied. The muid usually equalled twelve setiers or twenty-four mines,[4] but this is no more than a relationship between unknown quantities until the specific setier or muid has been converted into a standard and comparable unit of measurement. Fortunately, there is extant a document from the early four-

1. Fawtier, Hist. Gen., p. 202.

2. Dupont-Ferrier, Etudes, Vol. 1, p. 19.

3. Labande, "L'administration du duc d'Anjou," p. 7.

4. L. Delisle, Etudes sur la condition de la classe agricole et l'état de l'agriculture en Normandie au Moyen Age, Paris: H. Champion, 1903, p. 544.

teenth century which compares the measures of many French towns to that of Paris.[1] This enables us to convert measures, otherwise meaningless, into setiers of Paris and from there into modern hectolitres, since the setier of the latter city was equivalent to 1.55 hectolitres. Delisle has prepared, for Normandy, a list of the variations in the sub-multiples of the modius or muid by the towns where they were utilized.[2] Using these two sources of information, I have, where possible, converted the measures into hectolitres and given the price per hectolitre. Where it has not been possible to do this, the price is, of course, useless and has been discarded. Since this operation has already been performed in many of the sources, it has only been necessary to undertake this conversion for the Norman towns. The most important towns and measures are: Paris, 1.55 hectolitres the setier; Chartres, 1.27 hectolitres the setier; Arras, 3.51 hectolitres the setier; Orleans, .3361 hectolitres the mine; and Rouen, 1.9696 hectolitres the mine. The reader is not asked to believe that these measures are really accurate to four decimal places but they have been retained first because it seems no more accurate to omit them and second because they are the norm of the metric system, representing hectolitres, litres, and centilitres.

In England, the grain measure is the quarter of eight Winchester bushels, each composed of eight gallons. I have accepted this as a standard measure, following N. S. B. Gras,[3] but only in Oxfordshire and the immediately surrounding area. The uniformity of measures over England has been questioned by W. H. Beveridge,[4] who has found mention of quarters of from eight to ten bushels in some of the Winchester manors. D. L. Farmer[5] has, I think, presented a cogent case against this view for the period prior to 1318 in arguing that the price dispersion and the impediments to transfer of grain between manors by measure which one would expect to arise from variations in the measures do not appear. He does not, however, dispute Beveridge's findings for the period around 1354. This, fortunately, need not damage our assumption of uniformity since Beveridge, himself, in speaking of the districts from which we have drawn price data states that "Oxford, Berks, Bucks and Surrey are without mention of special measures in 1318 as they are in 1354."[6]

The Winchester corn-bushel, designated as the standard of the earlier system by Henry VII in 1495, still exists and it contains 2, 150.42 cubic

---

1. Published in Domino Du Cango, Glossarium Mediae et Infimae Latinatis, Niort: L. Favre, 1885, Vol. 5, pp. 436-8.

2. Delisle, Etudes, pp. 539-64.

3. Gras, Evolution of the English Corn Market, p. 132.

4. W. H. Beveridge, "Wheat Measures...," Idem, "The Yield and Price of Corn in the Middle Ages," Economic History, Vol. I, 1926-9, pp. 156-57; Idem, "A Statistical Crime of the Seventeenth Century," Journal of Economic and Business History, Vol. I, pp. 503-33.

5. D. L. Farmer, "Some Grain Price Movements in Thirteenth Century England," Economic History Review, December, 1957, pp. 207-20.

6. Beveridge, "Wheat Measures...," Idem.

inches.[1] The quarter composed of eight such bushels would be 9.955 cubic feet or 2.82 hectrolitres.

Enough has been said, perhaps, of the reservations which must be held in mind when observing price series from the two countries to allow their presentation in Graph III-2. The English series is in shillings per quarter; the French is in livres per hectolitre. In both cases the prices are simply expressed in the monetary units in which they existed with no reference to the metal content of the unit of account.

With the exception of some of the more obvious peaks such as the famine of 1315-17 and the famine following the advent of the Black Death of 1348-49, the similarity of the two series is far from obvious. The most striking movement on the French price graph, that of 1359-60, is entirely absent from the English series, as, indeed, might be expected in the year of maximum debasement in France.

Despite the apparent divergence between the two series, some similarity, however, does remain. In the period from 1365 to 1395 both economies enjoyed essentially stable money. English mints drew 300 pence from the pound Tower throughout the period so that changes in English money may be eliminated as a cause for price movements. In France there were two permanent but relatively minor debasements, one in 1384 which drew 6.25 livres from the mark as against 6.00 previously, and one in 1389 which made the mark worth 6.75 livres. In 1370 one mint, St. Pourcain, struck 85,000 ten denier pieces at 13.20 livres to the mark.[2] This was for one issue only and its importance is reduced by the fact that it occurred in isolation in a period of stability when all other mints drew 6.00 livres from the mark.

In Graph III-3, I have plotted the price of wheat during this sound money period. The right hand scale is livres per hectolitre and refers to the solid black line of French prices. The left hand scale applies to the English or dotted line and is in shillings per quarter. There is no difference between this graph and Graph III-2 except that the French scale has been doubled in order to make it visually more comparable with the English scale. The similarity of the two price lines is, I think, adequate to justify my contention that, in the absence of monetary change, the English and French series are close enough to provide a control for the degree of adjustment to debasement. If, after both French and English prices have been expressed in constant money, the two series remain as similar over the century, it does not seem unreasonable to conclude that debas ement was accounted for through price increases.

The two places on this graph, 1369-71 and 1391-92, where the congruence seems weakest may be discounted as calendar variations. At the beginning of this section, it will be recalled, the error which might arise from this source was seen to be as much as twenty-two months. We are, as a result, I think, justified in assuming the coincidence of the peaks during the years at hand. Both becaus e the volume of grain trading between France and England was limited, as discussed in the previous section, and because of the pattern of the graph itself, international trade cannot be a

1. H. W. Chisholm, On the Science of Weighing and Measuring and Standards of Measure and Weight, London: Macmillan, 1877, p. 65; Wm. Hallock and H. T. Wade, Outlines of the Evolution of Weights and Measures and the Metric Sys tem, New York: Macmillan, 1906, p. 35.

2. de Saulcy, Recueil, p. 512.

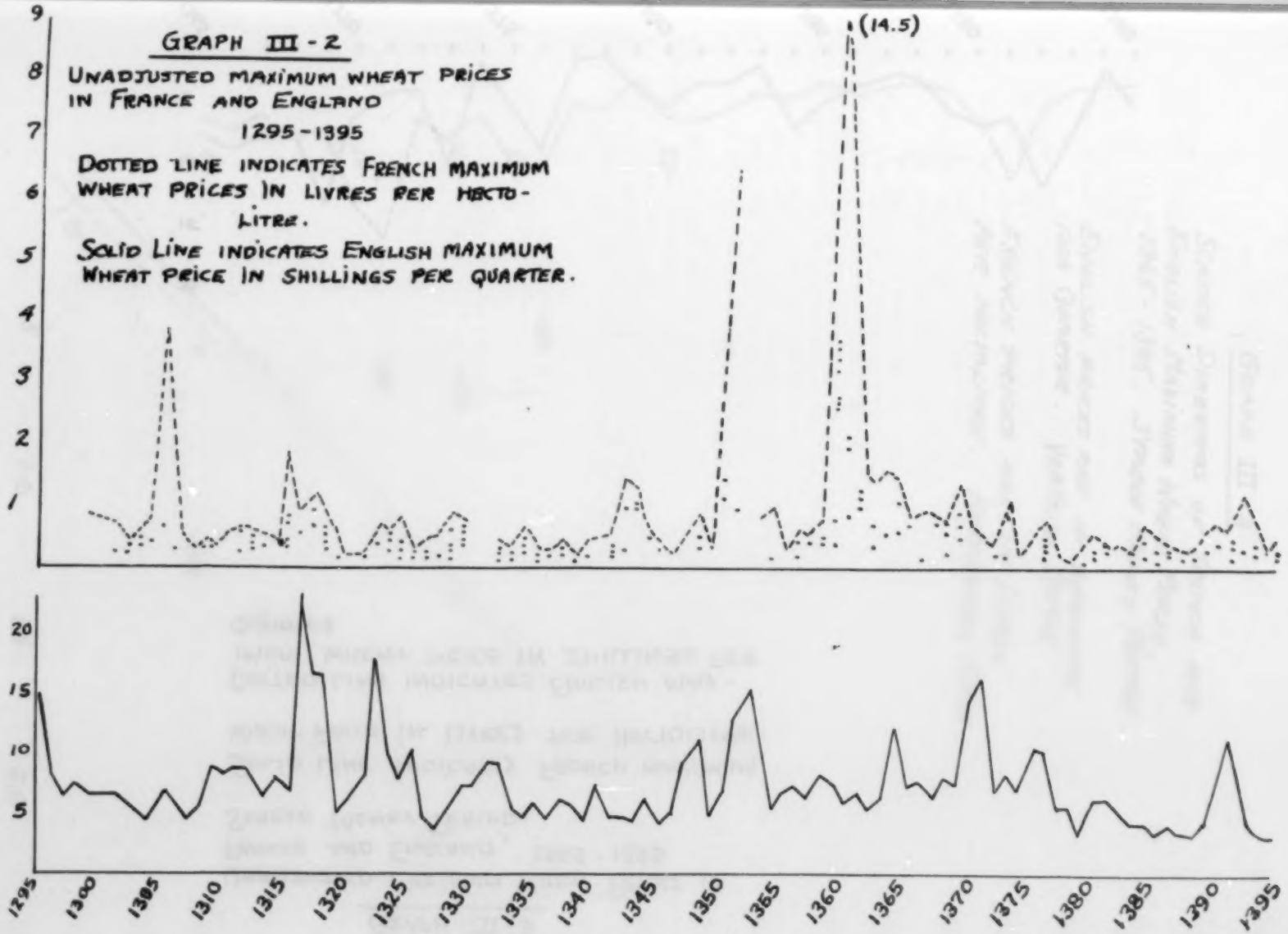
GRAPH III - 2

UNADJUSTED MAXIMUM WHEAT PRICES IN FRANCE AND ENGLAND

1295 - 1395

DOTTED LINE INDICATES FRENCH MAXIMUM WHEAT PRICES IN LIVRES PER HECTO-LITRE.

SOLID LINE INDICATES ENGLISH MAXIMUM WHEAT PRICE IN SHILLINGS PER QUARTER.

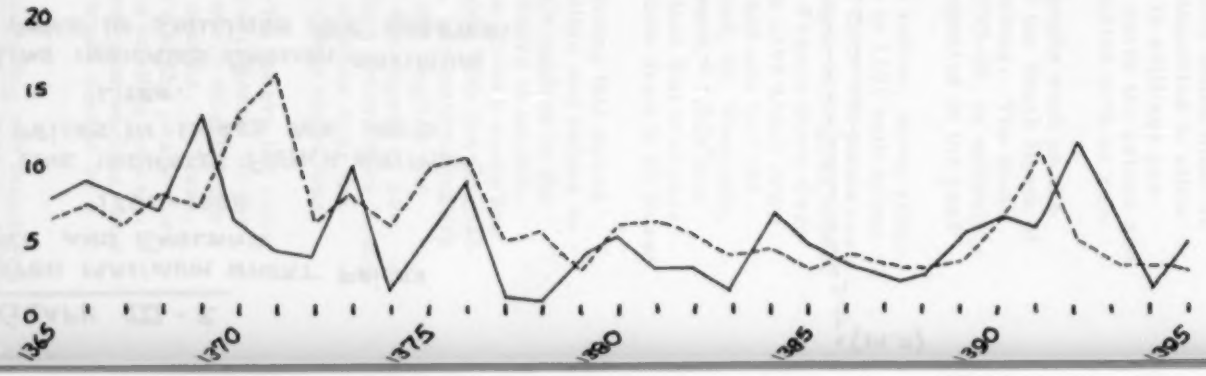


GRAPH III-3

UNADJUSTED MAXIMUM WHEAT PRICES IN FRANCE AND ENGLAND, 1365-1395. STABLE MONEY PERIOD.

SOLID LINE INDICATES FRENCH MAXIMUM WHEAT PRICE IN LIVRES PER HECTOLITRE.

DOTTED LINE INDICATES ENGLISH MAXIMUM WHEAT PRICE IN SHILLINGS PER QUARTER.

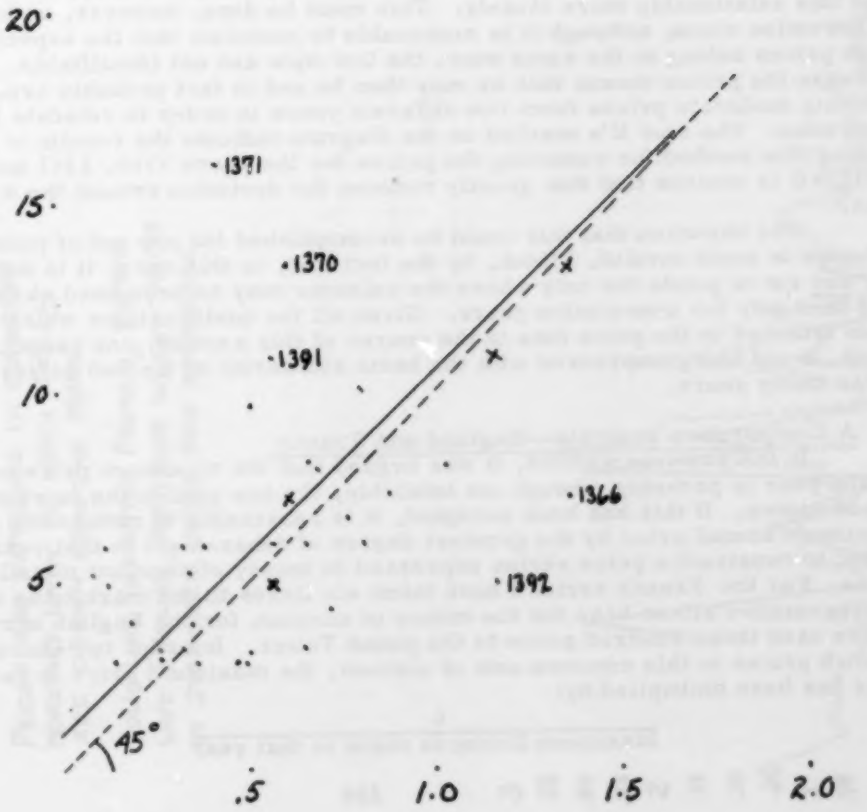


GRAPH III - 4

SCATTER DIAGRAMS OF FRENCH AND ENGLISH MAXIMUM WHEAT PRICES, 1365 - 1395. STABLE MONEY PERIOD.

ENGLISH PRICES ARE IN SHILLINGS PER QUARTER. VERTICAL SCALE

FRENCH PRICES ARE IN LIVRES PER HECTOLITRE. HORIZONTAL SCALE.



mechanism for exporting, through increased demand, the high price in one country after a lag of one year. Clearly, the famine price of 1369 in France could not have been lowered through imports of English wheat on a scale grand enough to produce a famine in England. This becomes an even more unlikely possibility when one remembers that in 1370 an English army was passing through France on a raid originating in Calais, passing Troyes and Noyon and then crossing France to Bretagne.

In the scatter diagram, Graph III-4, the prices are again plotted with England on the vertical axis and France on the horizontal axis. If the prices were in comparable units of measure and money, common movement of the two series would be indicated by a series of points along the dotted 45° line. If the units differed by some constant, as, for example, if the English quarter were larger than a hectolitre and by more than the simple adjustment of the scales compensates for, the points would, in the event of common movement, group around some line parallel to the dotted one. The solid black line is an estimated regression line for these points and it clearly is quite close to what might be expected if the two price series were congruent. The wide deviation about the line arises in part from the calendar variation. It is to some extent self-compensating in this diagram because of the sharp nature of the price movements. This is the case because the situation where a very high price in one country is matched with a low one from the other will tend, if the only reason for this is the calendar difficulty, to be the reverse stiuation in the following year. I have labeled the points for 1369-71 and 1391-92 in order to show this.

By simply reversing the prices for these years, one may, perhaps, see this relationship more closely. This must be done, however, with great reservation since, although it is reasonable to postulate that the especially high prices belong in the same year, the low ones are not identifiable. To reverse the prices means that we may then be and in fact probably are, plotting moderate prices from two different years in order to relocate the high ones. The four X's marked on the diagram indicate the results of following this method for removing the prices for the years 1369, 1391 and 1392. It is obvious that this greatly reduces the deviation around the 45° line.

The objection that this could be accomplished for any set of points so situated is made invalid, I think, by the fact that, in this case, it is not done for any set of points but only where the calendar may be presumed at fault and then only for consecutive years. Given all the qualifications which have been attached to the price data in the course of this section, one cannot, I think, avoid being impressed with the basic similarity of the two series for these thirty years.

4. A Comparative Analysis--England and France

In the previous section, it was argued that the maximum price paid in any year is probably, though not infallibly, the one paid in the most debased money. If this has been accepted, it is reasonable to reduce the maximum annual price by the greatest degree of debasement in that year in order to construct a price series expressed in money of constant metallic value. For the French series I have taken six livres to the mark to be a representative silver base for the money of account; for the English series I have used three hundred pence to the pound Tower. In order to reduce French prices to this common unit of account, the maximum price in each year has been multiplied by:

$$\frac{6}{\text{Maximum livres to mark in that year}}$$

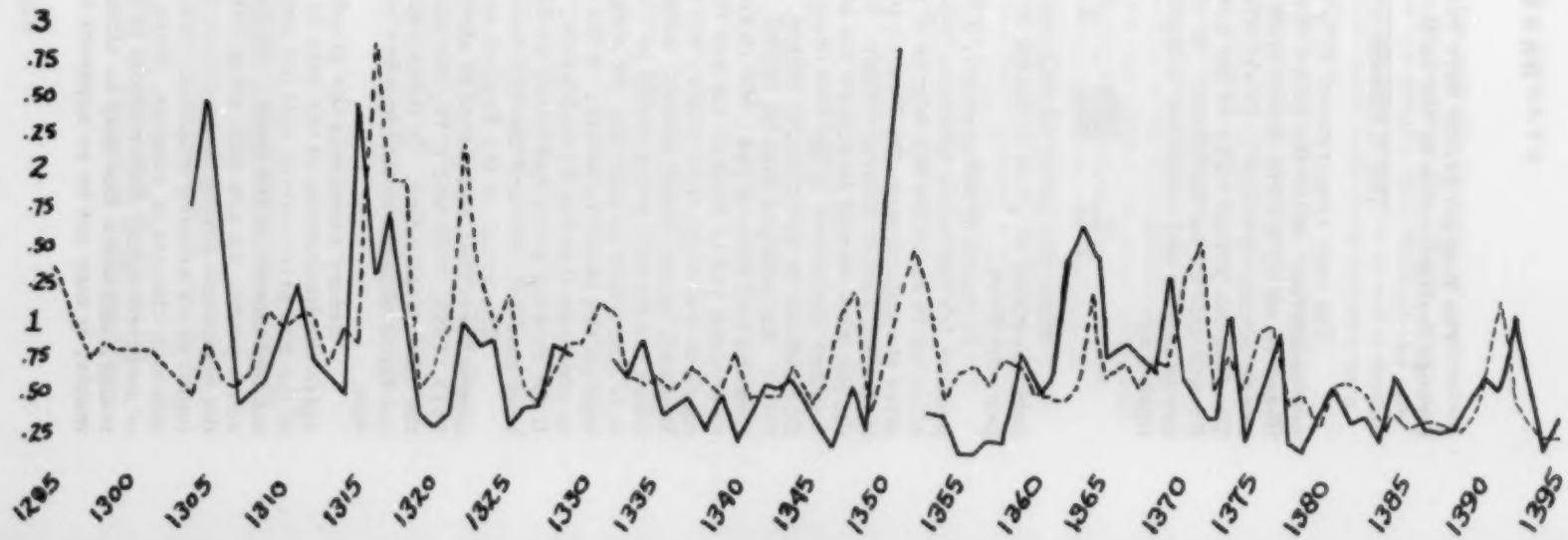
GRAPH IV-1

GRAPH IV-1

ADJUSTED MAXIMUM WHEAT PRICES IN FRANCE AND ENGLAND, 1295 - 1395

PRICES ARE EXPRESSED IN STANDARD MONETARY UNITS OF WHICH SIX (6) ARE DRAWN FROM THE MARK.

SOLID LINE INDICATES FRENCH, DOTTED LINE ENGLISH, MAXIMUM WHEAT PRICES IN STANDARD UNITS PER HECTOLITRE.



The English prices have been expressed in a common unit of account through multiplication by the factor:

$$\frac{300}{(20) \text{ X (Maximum pence to lb. Tower in that year)}}$$

The next requirement is to make the units comparable between the two countries. After the above operations, French prices are expressed in livres of which six are drawn from the mark of 4,608 grains; English prices are in pounds of which 1.25 are drawn from the pound Tower of 5,400 grains. The English prices refer to the quarter of 2.82 hectolitres whereas the French unit is the hectolitre. In order to convert English prices so that they are in money and measures comparable to French prices each must be multiplied by:

$$\frac{6}{1.25} \cdot \frac{5400}{4608} \cdot \frac{1}{2.82} = 1.99454$$

When this latter multiplication is completed the price of wheat in both countries is given in terms of livres or pounds of identical metal content per hectolitre.

In years of debasement, French prices have been reduced to compensate for the lower intrinsic value of the money of account. This fact enables us to measure the degree of price adjustment to monetary change. If, after the adjustment, the French price is significantly below the English or if the two series diverge widely, it is presumably because the French price did not rise enough to absorb the alteration in the money. The similarity of weather conditions in the two countries allows us to some extent to ignore fluctuations in agricultural output.

An example may be useful. If the unadjusted price of wheat in both England and France was .500 livres per hectolitre in each of two consecutive years and in France six and twenty-four livres were drawn from the mark in the respective years, the adjusted series would diverge widely. England, under sound money, would retain the price of .500 in both years while the French price would be .500 in the first year and .125 in the second. It is important to note that the slopes of the two graphs would diverge as well as the absolute levels. If the data were perfect, we would be compelled to conclude that the French price, in this case, did not adjust to debasement. If the French price had been, in the second year 2.000, the adjustment would have made it .500 and the conclusion would be just the opposite.

The graph of the English and French price series after these adjustments have been completed is shown in Graph IV-1. In six years, 1311, 1323, 1333, 1353, 1355 and 1370, the maximum debasement is known to occur after the date of the price. In these years the adjustment has been made in accordance with the actual number of livres drawn from the mark at the price date.

Before examining the graph more closely, it will be useful to review again the limitations of the data so that the very similarity and near congruency of the two price series will not tempt us to expect too much in interpreting the movements of the lines. No matter how close the relationship seems in some years, we are still using only one price in each year, and while this is the maximum among those prices which remain to us, it may not be the maximum price actually attained. We are comparing two distinct countries and although climate is similar, there surely may be variations in the intensity of meteorological disturbances in the two. The uncertainty of price dates remains, and while this may be allowed for in years of exceptional price movements, it may not be so apparent in an average year.

Because of these calendar difficulties and also because of qualitative differences in the data from year to year, the analysis is, I think, more appropriately conducted through inspection of the graphs than through the use of correction coefficients. It is not, in my opinion, possible to set a meaningful significance level for a statistical test based upon such diverse data as presently exist for fourteenth century France.

In short, enough flaws exist in the data to account for as much noncoincidence of movement as is visible on the graph in most years. It is not necessary to conclude that debasement reduced the real French price level whenever the French series falls below the English series: even for the years of greatest disparity there are often factors, external to monetary change, which may account for the diversity of movement in the two graphs.

In order to set some limit upon the size of the incongruencies which may be ascribed to faults in the material we may take a few years of monetary stability and examine the degree of variation between the two series. Since the precondition for the selection of these years is the absence of monetary change, the two series can diverge only because of problems inherent in the data or in the method. During the periods 1374-78 and 1381-83, six livres were drawn from the mark and had been since 1364 with the exception of a minor output of debased coin in 1370.[1] Reference to these years during the examination of the graph will, I think, provide the reader with some estimate of the variation which might be considered normal to the series in the absence of debasement.

Although there is a price peak in both countries in 1305, indicating that similar climatic conditions are involved, the French series rises far above the English during the years 1304-5. In northern France, however, there is record of a "grande dissette."[2] Philippe le Bel made an attempt to set a maximum price for wheat but this did not succeed. The royal agents may have contributed to the scarcity by aiding in the illegal export of wheat from Paris to the provinces in order to obtain a price higher than the maximum of Paris.[3] Since there is a price rise but no record of comparable shortage conditions in England, the difference between the price levels in the two countries is probably a qualitative one, magnified by the process of selecting the highest available price in each year. Although the known difference in the relative scarcity between the two countries forbids precision, it seems reasonable to conclude that the French price rise was adequate adjustment to the fact that 9.20 livres were drawn from the mark in 1305 compared to 3.01 in 1306. The evidence is certainly not inconsistent with this statement.

Shortage conditions seem to have been more intense in England than in France in the period 1320-25. Miss Robbins finds evidence of semi-famine conditions in England in 1321 but makes no such comment on France for this year in her comparison of the two.[4] The pattern of the two price

---

1. Supra, Section 3.

2. E. Levasseur, Le prix aperçu de l'histoire économique de la valeur et du revenue de la terre en France (Paris, 1893), p. 111.

3. E. Boutaric, La France sous Philippe le Bel (Paris, 1861), pp. 364-5.

4. H. Robbins, "A Comparison of the Effects of the Black Death on the Economic Organization of France and England," Journal of Political Economy, Vol. 36, 1928, pp. 447-479.

series in these years is, however, essentially the same, particularly if one places the French peak of 1324 in 1325. Since this peak is, in fact, an undated price from Cani and since there is evidence of a drought in both countries in 1325,[1] this seems a permissable adjustment.

A specific examination of the monetary changes during these years will aid in the evaluation of the adjustment of price to debasement. In 1320, 1321 and 1322, 3.86 livres were drawn from the mark in France. There was no alteration in these years and stability of the coinage had been continuous since 1312. A monetary explanation founded simply upon debasement cannot be adequate for the large excess of the English price series over the French series during these years. When, in 1323, debasement takes place and 5.43 livres are drawn from the mark, 4.52 for our price, the two series converge more closely than in the immediately preceding years of stable money.

In 1324 and 1325, the number of livres to the mark remains 5.43 but the two series seem to vary widely unless the proposed coincidence of the French 1324 price and the English 1325 price is accepted. If this is done, the apparent divergence of the two series is so reduced as to require no further comment. If this adjustment is not accepted, an explanation of the variance must be sought elsewhere than in debasement. No monetary change occurs between 1324 and 1325 to account at once for the near identity of the two series in 1324 and the deviation in 1325. In either case the fluctuations of the two graphs throughout the period, 1320-26, cannot be taken to be the result of mutations of the coinage.

In 1340 the number of livres to the mark increases from 6.00 to 10.50. The two series diverge sharply and by more than may be accounted for by shifting the calendar. Even if the price from 1339 were substituted for that of 1340, it would, after the adjustment for the change in the unit of account, be almost as low as the one which is used in this year. The ever present chance that the price was paid before the debasement or in better money could be responsible for the spread, but, in the absence of other evidence, it must be acknowledged that it might result from a failure of prices to rise enough to compensate for the decreased intrinsic value of the unit of account.

After this year, two further periods, 1346-48 and 1355-58 would seem to require specific comment. In the former, the French price is consistently low although there is only one currency mutation which affects the maximum number of livres drawn from the mark. In 1347, this maximum increases from 6.11 to 9.00. Despite this change in the French unit of account, the slopes of the two graphs remain similar. The evidence is here inconclusive, and would support either a positive or a negative statement regarding the ability of French prices to compensate for debasement. In absolute terms, the difference between the two series is greater after the debasement than before it. In relative terms, the French price in 1346 before debasement is only 21.9 per cent of the English price whereas, after the debasement, it is 47.8 per cent of the English price. While it is evident that it is not possible to conclude from these figures that prices rose in proportion to debasement, it is equally apparent that there are no grounds for the opposite conclusion.

During the years 1355-58, the French series again falls well below the English in a period where the number and magnitude of the changes in the money is quite significant. The separation of the two price lines is here, I think, attributable to the paucity of the data, rather than to any economic

1. H. Robbins, op. cit., pp. 453, 459.

response or lack of response to coinage mutation. For the entire four year span, we have only nine prices from which to select the maximums needed to measure the response to debasement in a period which includes twenty-five mutations of the coinage. In 1355 there is only one price, and in 1357 there are only two. In 1355 where both six and eight livres were drawn from the mark at the price date, the wide differential between the two series cannot be explained by the failure of prices to rise in compensation for debasement. The price was paid at a time when the number of livres to the mark was close to a five year low. For the remainder of the years in this period, 1355-58, it is necessary to admit that there is simply too little information to arrive at any valid conclusion. It is, perhaps, better to ignore this portion of the graph, rather than to endeavor to account for its behaviour in monetary terms. If, for example, as in 1356, 6.00, 8.00, 12.00, 15.00 and 30.00 livres were drawn from the mark, the chance of one of the three prices for this year having been paid in the most debased coinage is small. Even if a proportional price adjustment took place as twelve instead of six livres were taken from the mark, our process of adjustment, predicated on the assumption that the maximum price coincides with the maximum debasement, in this case 30.00 livres to the mark, would reduce the price based on money of 12.00 livres to the mark to a point far below English levels.

Despite the fact that there are many areas where the conclusions which may be drawn from the two graphs are clouded, there are many segments where the two lines approach each other so closely as to be virtually identical. This is, for example, the case between 1386 and 1390 under essentially sound French money. It is also true in years of major debasement, where the probability is smaller, owing to the chance that a price paid in better money will erroneously be reduced by the maximum debasement, that the method of reduction will be truly accurate.

In 1304-5, considerable debasement occurred and real French prices rose so far above the English level that we have been compelled to seek an explanation in the relative intensity of famine conditions rather than in monetary change. In 1311, a year of roughly thirty per cent debasement, from 3.01 to 3.93 livres to the mark, the French price is slightly above the English, and both series remain extremely close to each other from 1310 to 1313. These facts indicate that the alteration in the money was fully absorbed through price flexibility.

From 1326 through 1329, when the French graph is temporarily interrupted, the two series again almost overlap, despite the fact that the maximum number of livres drawn from the mark in these years is respectively 6.00, 6.00, 6.31 and 6.00. For this period, we may, I think, again conclude that price was sensitive enough to make the adjustment required by monetary changes. The series are closer in this period than in the immediately preceding one, 1320-25, when 3.86 and 5.43 livres were drawn from the mark. The similarity compares very favorably with the reference years 1374-8 and 1381-3.

The segment from 1334 to 1340 yields no firm evidence of either a positive or negative character. From 1341 through 1345, however, the two graphs are not merely parallel but are virtually coincident. During this period the annual maximum numbers of livres to the mark are: 1341, 12.00; 1342, 15.00; 1343, 13.75; 1344-45, 6.11. For the latter four years, the adjustment is particularly impressive since the range of monetary change, from 6.11 to 15.00 livres to the mark, which had to be absorbed in price changes, is far larger than that for the less conclusive years for 1334 to 1340. The range of the debasement in the latter period extended only between 4.15 and 6.00 livres to the mark.

The ability of the French price level to absorb even very large mutations of the coinage may be observed in the years 1359-61. Excluding the various levels below the annual maximums, the French king, in these years, drew respectively 45.00, 125.00, and 18.00 livres from the silver mark. Comparison of the adjusted price levels in France and England during these years reveals as great similarity as is evident in any other period throughout the century.

After these years, the money is stable, except for three minor debasements, one in 1364, one in 1384 and one in 1389. In each of these years, the French and English series are similar enough to provide support for the assertion that price adjusted to debasement in these years also.

The discussion of the price data undertaken in this section has been predicated upon theoretical considerations derived from the first and third sections. In the first chapter of the dissertation, it was argued that many economic, demographic, technological and political variables were, broadly speaking, similar and co-temporal in France and in England over the course of the fourteenth century. In the third section, we endeavored to select an area of France which might reasonably be expected to have almost identical weather conditions to England. Finally, accepting the findings of the first chapter of the dissertation, we were led to the conclusion that wheat prices, motivated by similar climatic conditions should follow a parallel course in the absence of monetary change. It was shown in the beginning of this section that, if prices rose in proportion to debasement, reduction of prices by the debasement would remove the effect of monetary change and again lead us to expect the two series to follow a parallel course. This was done, resolving the problem of the money of account in accordance with the evidence from the abbey of St. Denis, but only at the expense of a significant portion of the price data. The simultaneous existence of many royal coinages was proved and offered as an explanation of the wide range of prices in the years of debasement and also as the rationale for the exclusion of all but the maximum price in any year. The adjusted price graphs of the two countries were then used as a device to measure the response of French prices to debasement.

The general and consistent proximity of the English and French price graphs may, I think, alleviate some of the doubts which the reader may have regarding the historical validity of the similarity between the two countries claimed in the first section.1 Some explanation has been offered for the more pronounced disparities between the two series, but the final conclusion to be drawn with regard to French price flexibility in relation to debasement remains in some years of interpretation rather than proof. In eighteen years, 1310-13, 1326-29, 1342-45, 1359-61, 1364, 1384 and 1389, it is not possible, I think, to doubt that the French price series made the precise adjustment required for the maintenance of the intrinsic value of the money of account. These years exhibit no common characteristic which might be expected to make them behave differently than any other years of the century. Among them are years of peace and war, of famine and plenty, and of French military success and failure. They are drawn from all parts of the century and include the period prior to the Black Death as well as the period after it. In short, there is no a priori reason to expect French prices to be more flexible in these years than in any other group of years.

Between 1310 and 1390, seventy-seven years and thirty-two changes in the maximum number of livres to the mark are represented. Fourteen of these changes fall in the eighteen years of almost perfect adjustment. In many of the other years it has, I think, been shown that it is reasonable to

conclude that the adjustment was made. In no year is it possible to conclude that the adjustment was not made. On the basis of the evidence, it would seem more reasonable to accept the hypothesis that newly contracted prices rose in proportion to debas ement than to deny it.

Summary: The purpose of this essay is to estimate the impact of tightening credit on total capital expenditures of United States municipalities which have planned or authorized the financing of some capital expenditures by long term borrowing.

A cross section of data on state and local government bond interest rates and municipal capital expenditures drawn from the 1955-57 period of tightening credit is analyzed to ascertain (1) which types of bond become relatively more expensive to issue when credit tightens, and (2) whether those municipalities which have planned or authorized the financing of capital expenditures by bonds which have become relatively more expensive to issue postpone and cut back a larger proportion of total capital expenditures than other municipalities.

If the increase in the interest rate is measured in absolute basis points, the evidence is that long term low rated bonds become relatively more expensive to issue when credit tightens. If the increase is measured in relative terms, the evidence suggests that low risk bonds become relatively more expensive, with one exception. When credit tightens, the percentage increase in the interest rate on long term bonds is greater than the percentage increase on short term bonds.

The evidence indicates that an increase in borrowing costs has a negligible impact on long range capital improvement plans for total capital expenditures. However it supports the hypothesis that officials respond to an increase in borrowing costs by postponing and cutting back the total volume of actual relative to authorized capital expenditures.

This essay was submitted to Yale University in 1961 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently an assistant in research, Cowles Foundation for Research in Economics at Yale University.

Dissertation Committee: Professors Challis A. Hall (Chairman), Hugh T. Patrick, and Harold W. Watts.

*

I wish to thank Professors Challis A. Hall, Arthur M. Okun, and Harold W. Watts for their valuable guidance in the preparation of the dissertation from which this essay is taken. I am also grateful to the Cowles Foundation for Research in Economics at Yale University for financing the computations for the multiple regression analysis and the Commission on Money and Credit for financing the research during the 1959-60 academic year. The results of this research were incorporated in a monograph for the Commission.

C. D. P.

THE IMPACT OF TIGHTENING CREDIT ON MUNICIPAL CAPITAL EXPENDITURES IN THE UNITED STATES

Charlotte DeMonte Phelps

1. A Framework for Analysis

The purpose of this essay is to estimate the impact of tightening credit on total capital expenditures in municipalities which have planned or authorized the financing of some capital expenditures by long term borrowing.[1] The question is apt because municipalities finance almost half their capital expenditures by net increases in debt.

A. The Decision Process Which Controls the Planning and Execution of Municipal Capital Expenditures

To analyze how tightening credit may affect the volume and timing of municipal capital expenditures, it is useful to date the period for which expenditures are planned and to divide the decision process which controls the planning and execution of plans into four stages: (1) the initial formulation of expenditure plans for t_1 and t_2, (2) the authorization of expenditures planned for t_1, (3) the review and execution of expenditures authorized for t_1, and (4) the review and revision of expenditures planned for t_2.[2] Stages (1) and (2) occur within one fiscal year which may be designated t_0. Stages (3) and (4) occur within one fiscal year--t_1. The duration of stages (1), (2) and (4) is assumed to be three months.

Suppose that in t_0 municipal officials plan capital expenditures for periods t_1 and t_2. For various reasons, they generally prefer to plan finances so that the tax rate changes gradually or not at all over the duration of the plan. Such a policy produces a smooth revenue stream which may be either stationary, rising, or falling depending upon what changes take place in the tax base. Since capital projects are "lumpy," it is necessary either to save in advance or to borrow to finance expenditures which cannot be timed to match the smooth revenue stream. The same need to borrow or save arises if the desired stream of revenues does not happen to conform with the preferred stream of expenditures. Thus interest enters plans either as the cost of borrowing or the return from investing savings.

Interest makes the cost of any capital project depend on the timing of the planned expenditures. The benefits from the project services may also depend on their timing, so that the problem of reconciling the schedule of expenditures with the schedule of revenues may be a complicated one. Fortunately, these complications do not exclude the possibility of a few generalizations concerning the effect of the interest rate and interest rate changes on the volume and timing of planned expenditures.

Suppose that officials anticipate a certain time path of the interest rate over the planning horizon. For simplicity assume that they expect the interest rate in t_0 to prevail in future periods. If the interest rate in t_0 is high, officials will wish to borrow sparingly and to defer many outlays until they have accumulated reserves to finance them. The opposite is true if the interest rate is low. A high interest rate also means that the volume

---

1. Total capital expenditures are defined to include outlays for land, equipment, existing structures and construction financed from current tax revenues and long term borrowing.

2. The definition of the period "t" is adaptable to any specified planning period. It is convenient to think of the period for which plans are made as a fiscal year.

of capital improvements which they can purchase with any given budget is smaller than with a low interest rate.

In t_0 officials submit plans for t_1 to the city council to be authorized (approved), but they leave plans for t_2 tentative.

In t_1 officials review plans in the light of conditions which have developed since the initial formulation of plans. Changes in priorities of proposed projects, changes in expected tax revenues and changes in credit conditions, for example, may prompt them to make revisions.

If credit in t_1 is tighter than it was in t_0, officials review the courses of action open to them if they are to carry out expenditures in t_1 as authorized. Assuming that some bond financed expenditures have been authorized, they can either sell the authorized bond as planned or postpone the bond sale and secure temporary finance to make necessary cash outlays. Which, if any, of these courses is feasible depends on five factors: (1) legal restrictions on the bond sale, (2) the government's policy of timing the bond sale with respect to the scheduled initiation of construction, (3) arrangements for the substitution of revenues within and between separate Fund accounts, (4) limitations on short term borrowing, and (5) restrictions on increases in revenues to cover additional interest payments.[1]

The most common legal restrictions leading to bond sale postponements are interest rate[2] and tax rate specifications in the bond ordinance. Although these restrictions can be amended, the amendment process takes time and does not remove the necessity of postponing the bond sale. General statutory limitations on the maturity of bond issues also cause bond sale postponements by making it impossible to stay within the budget constraint by lengthening the amortization period.

The government's policy of timing the bond sale with respect to the scheduled initiation of construction affects its freedom to postpone the bond sale without having to postpone expenditures. If officials have planned the sale considerably in advance of the time they expect to be ready to make expenditures, they can postpone it without delaying construction. But if they have planned the sale to coincide with the time they expect to make expenditures, they can postpone the sale and carry on the expenditure program only if they secure temporary finance.

If officials have planned to sell the bonds while construction is in progress, they will already have been using short term finance to make progress payments on construction. Their freedom to carry on construction while postponing the sale will depend on their ability to continue to use temporary finance. Of course if construction has been completed, officials can postpone the sale without affecting the rate of construction.[3]

---

1. Municipalities usually keep separate accounts for general tax revenues, bond sale proceeds, and revenues from special sources.

2. There are also general statutory rate limitations. In recent years they have not necessitated bond sale postponements, because they are above recent market rates. See Dick Netzer, "State-local Response to Changing Credit Conditions: The Institutional Obstacles," Journal of Finance, XV (May 1960), pp. 224-225. See also Table 12 in this essay.

3. Officials must follow the policy of selling the bond before beginning construction if the city charter requires the sale before permitting obligations to be charged against the bond financed expenditure authorization. For an indication of the number of cities which have such a requirement, see Table 12.

Consider now the various sources of short term finance which might be available to carry on the expenditure program in the event that officials were to decide to postpone the bond sale. Broadly speaking, there are two kinds of temporary finance: internal and external. The internal sources consist of the cash balances in the various government Fund accounts. Depending on the restrictions governing the temporary substitutability of authorized funds for authorized purposes, officials can draw down these cash balances to meet disbursements. In some governments, statutes make possible substitutions between the proceeds of bond sales by permitting a consolidated Bond Fund--a fund in which disbursements for one authorized bond financed project can be met out of the proceeds of another bond sale--or by permitting the sale of multiple purpose bonds.[1] In some governments, statutes permit temporary substitutions between other Funds (general revenues) and Bond Funds. In other words, disbursements from Bond Funds can be met out of a loan from the General Fund if the finance officer wishes to delay selling the bond. When such arrangements have been made, officials do not have to relate each bond sale to the expected stream of expenditures charged against its authorization. Rather, their timing policy can be oriented toward maintaining an adequate cash balance to meet obligations for all purposes combined.

External temporary finance may take the form of borrowing short term from a bank or floating a short term note or bond anticipation note on the open market. The feasibility of external short term finance depends on state and local laws and customs. The Federal Reserve Bank survey of municipal financing in 1956 found a wide range of state and local short term borrowing practices.[2]

Laws which prohibit short term borrowing--and taking a less extreme example--laws which require a down payment on short term loans are deterrents to short term borrowing. On the positive side are laws which permit the government to use short term borrowing a relatively long time before requiring permanent funding. It is in places operating under such laws that it is customary to finance bond expenditure authorizations from short term loans until after the project is completed and exact costs are known.

With respect to the last of the five factors, general statutory limitations on the proportion of the tax rate to be devoted to interest payments are rare. However a common rule of thumb is that interest plus principal payments should not be greater than 25% of the budget.

There are legal and economic limitations on increases in project revenues to service the debt on so called "revenue" bonds. Restrictions on landing fees at airports are an example of a legal limitation on airport revenue bonds. Economic limitations come from an inelastic demand for the services of the project to be financed by the revenue bond. In the event that additional project revenues cannot be raised, officials sometimes pledge the taxing power of the municipality as additional collateral.

After considering the feasible alternatives, officials will ask whether the most acceptable of them--i.e., the least costly action--is preferable to

---

1. For an indication of the number of cities which have arrangements for internal substitutability of funds, see Table 12.

2. Ralph A. Young, "Recent Developments in State and Local Government Financing," Unpublished memorandum, Board of Governors of the Federal Reserve System, April 17, 1957.

postponing or cutting back the level of capital expenditures in t_1.[1] Notice that the choice is whether to postpone or cut back the level of total capital expenditures, not just the level of bond financed expenditures. The relative prices and priorities of the improvements determine which shall be postponed or cut back.

Whether officials prefer to carry out expenditures as planned depends in part on the proportion of capital expenditures they have planned to finance from bond sales. The larger is the proportion of planned bond financed expenditures, the greater is the increase in interest payments resulting from a given increase in the interest rate and the greater is the incentive to cut back or postpone expenditures.

The preferable course of action also depends on the magnitude of the change in the municipal bond rate between t_0 and t_1. The larger is the change, the greater is the incentive to cut back or postpone expenditures. It is not known whether a given increase in the interest rate will cause more postponements and cutbacks in a city which initially expected to pay a high interest rate than one which initially expected to pay a low rate of interest.

Last, the preferable course of action depends on officials' expectations concerning the level of the interest rate in t_2. If officials expect the interest rate in t_2 to rise above the rate in t_1, they will not postpone expenditures that were planned for t_1 to t_2. However, they may cut back the level of expenditures to be made in t_1. If they expect the rate in t_2 to fall below or remain at the level of t_1, they may either postpone or cut back the actual level of expenditures in t_1.

Officials also review plans for expenditures in t_2 in the light of tighter credit conditions. Whether officials prefer to leave the volume of expenditures planned for t_2 unchanged depends on the same factors which affect their preference to carry out, postpone, or cut back the actual level of expenditures in t_1; i.e., on the degree of reliance on debt finance, the size of the actual change in the interest rate between t_0 and t_1, and the expected level of the interest rate in t_2.

If officials decide to carry out some expenditures in t_1 as planned and they are unable or unwilling to raise taxes, they must cut back the volume of expenditures previously planned unless they expect the interest rate in t_2 to fall enough to compensate for the additional costs incurred by carrying out expenditures in t_1 as planned.

If officials expect the interest rate in t_2 to stay above the level of t_0 and they are unable or unwilling to raise taxes, they will have to cut back the volume of expenditures planned for t_1 and t_2 in the aggregate. Which projects they will drop depends upon their relative priorities and relative costs. If they expect credit to be tighter in t_2 than in t_1, they may attempt to accelerate some expenditures planned for t_2 to t_1.

If officials expect the interest rate in t_2 to return to the level of t_0 and if they decide to postpone their expenditures from t_1 to t_2, they will leave the volume of expenditures planned for t_2 in t_0 unchanged; that is to say, they will add the volume of postponements to the expenditures initially planned for t_2.

---

1. Special meanings are intended for the words "postpone" and "cut back." Expenditures which are postponed are included in plans for a future period. Expenditures which are cut back are withdrawn once and for all from the planning horizon.

B. Criticism of Approaches Used in Previous Studies

The discussion of previous work is confined to two studies of the impact of monetary policy on state and local government construction expenditures. The first was conducted by Frank E. Morris[1] and the second by Richard C. Pickering.[2]

Morris compared the cyclical movements of state and local government bond yields with the cyclical movements of the dollar value of actual bond sales, contracts awarded and construction put in place by state and local governments from 1952 through 1959. He found that the bond yields followed the contour of the general business cycle as defined by the National Bureau of Economic Research reference cycle dates, while the bond sales, contracts awarded, and construction put in place series showed a moderate contracyclical pattern. From the deviations above and below the trend line, he estimated that about 10% of bond sales, 5% of contracts awarded, and slightly less than 5% of construction put in place were shifted from the final stages of the boom to the recession and recovery periods.

Because Morris did not compare planned with actual expenditures, he failed to distinguish between postponements in expenditures due to tight money and advances in the timing of expenditures due to easy money. He lumped both postponements and advances together as shifts.

Pickering did not lump postponements and advances together. He measured the volume of construction expenditures which were postponed when credit tightened in the latter half of 1957 and the volume of construction expenditures which were accelerated when credit eased in the first half of 1958. The study was based on data obtained in a survey of state and local finance officers. The officials interviewed were selected on the basis that their cities or states issued bonds during the first half of 1958. They were asked when they had originally planned to sell the bond which put them in the sample. If they sold the bond later than they planned, a bond sale postponement was noted; if they sold the bond earlier than planned, a bond sale acceleration was noted. Additional questions were asked to determine whether and why officials did or did not delay (accelerate) expenditures when they postponed (accelerated) bond sales.

Pickering found that 9% of the bonds sold in the first half of 1958 were postponed from the preceding tight money period and 2% were sold in this period because of the easy credit conditions. He found that about 17% of the bond sale postponements resulted in postponements in construction and that none of the bond sale accelerations resulted in accelerated construction. To put his findings in Morris' terminology, he found that about 1.5% of construction was shifted from the final stages of the boom to the recession and recovery periods.

Pickering's method of sample selection and his measure of planned expenditures biased his estimates of postponements. By limiting the sample to governments which sold bonds in the easy money period, he excluded those which postponed projects in the tight money period and did not reschedule them when credit eased. By using data on planned bond sales to represent

---

1. Frank E. Morris, "The Impact of Monetary Policy on State and Local Governments: An Empirical Study," Journal of Finance, XV (May 1960), pp. 232-249.

2. Richard C. Pickering, "State and Local Government Bond Financing During the First Half of 1958." Unpublished manuscript which reflects the views of the author and not necessarily those of the Board of Governors of the Federal Reserve System, May 5, 1959.

planned bond financed expenditures he provided no way of estimating the volume of expenditures in his sample which were cut back before the bond sale was planned.

Both Morris' and Pickering's studies give the erroneous impression that the only effect of monetary policy is to shift the timing of expenditures from tight money periods to easy money periods. To estimate the volume of expenditures that are cut back when credit tightens and the volume of new projects that are scheduled when credit eases, it is necessary to get data on planned expenditures, not planned bond sales. By the time officials plan the actual date of a bond sale, engineering and architectural specifications for construction are generally underway, if not completed. Therefore postponement of a bond sale seldom means cancellation of a project. Officials are more likely to cut back plans at an earlier stage of the decision process--before expenditure plans are authorized.

Another limitation on both studies is their focus on construction expenditures. While it is true that the majority of long term bonds are used to finance construction, many governments have some arrangements for the substitution of revenues within and between separate Fund accounts. It is of interest to know whether they use these arrangements to adjust to changes in monetary policy. Do they, for example, borrow temporarily from appropriations for equipment purchases in order to carry out construction plans on schedule when credit tightens? Officials may prefer to vary the timing of cash purchases rather than the timing of contracts, because drawing up a contract involves a complicated multi-stage decision process.

Another facet of the adjustment process which neither Morris nor Pickering investigated is the extent to which current tax financed expenditures are postponed or cut back when credit tightens and accelerated or increased when credit eases. Arrangements for the substitution of revenues between the General Fund and Bond Funds open up the possibility of altering the timing of tax financed expenditures in order to carry out bond financed expenditures on time. In such cases officials will borrow some funds internally rather than on the open market in order to keep total expenditures over the planning horizon within the budget constraint. Even without these arrangements, it is possible that officials will prefer to cancel some tax financed authorizations in order to issue a bond as planned.

At the planning stage of the decision process, the method of finance is not legally tied to a particular project. Therefore officials may cushion the impact of tightening credit by switching some projects from debt finance to pay-as-you-go finance. This question is outside the scope of both previous studies because they did not investigate the impact of monetary policy on the planning of expenditures prior to their authorization.

C. The Approach Used In This Study

The approach used in this study is shaped by the conception of the governmental decision process described in Section A and by the fact that when credit tightens some governments experience larger increases in the cost of borrowing than others. A cross section of data on state and local government bond interest rates and municipal capital expenditures drawn from the 1955-57 period of tightening credit is analyzed to ascertain (1) which types of bonds become relatively more expensive to issue when credit tightens, and (2) whether those municipalities which have planned or authorized the financing of capital expenditures by bonds which have become relatively more expensive to issue postpone and cut back a larger proportion of total expenditures than other municipalities.

Part 2 is devoted to an analysis of the impact of tightening credit on

the cost of issuing different types of state and local government bonds. The bonds are typed according to seven characteristics which previous studies suggest define the structure of state and local government bond interest rates.[1] These characteristics, which may be viewed as measures of different kinds of risk, are: maturity, optional feature, credit rating, size of issue, type of collateral, type of issuing agency, and purpose of issue.

Two general hypotheses are tested by means of regression analysis. The first is that the structure of interest rates results from investors trading off risk against the rate of return, the consequence being that low risk bonds bear low interest rates and high risk bonds bear high interest rates. The second hypothesis is that the compensation investors demand for incurring a given risk varies with changes in monetary conditions. When credit tightens investors demand a higher rate of return on all state and local government bonds, but they demand a greater increase in the return on high risk bonds than on low risk bonds.

Two broad hypotheses are also tested in Part 3. The first is that the proportion of total planned capital expenditures that is postponed and cut back is positively related to the increase in borrowing costs experienced between the initial formulation of plans and the revision of plans. The second is that the proportion of total authorized capital expenditures which is postponed and cut back is positively related to the increase in borrowing costs experienced between the authorization and execution of expenditures.[2] Alternative measures of the increase in borrowing costs--the absolute increase and the relative increase--are used because it is not known whether a given absolute increase causes fewer postponements and cutbacks in a city which initially expected to pay a high interest rate than in one which expected to pay a low interest rate. The responses of current tax financed and bond financed expenditures are examined at both the planning and the execution stages of the decision process.

How are the hypotheses tested in Part 2 related to the hypotheses tested in Part 3? To test the hypotheses in Part 3, it is necessary to have estimates of the interest rates on which officials base planned, authorized, and actual expenditures. The results of the tests of the hypotheses in Part 2 lead to an equation which is used in Part 3 to provide the needed estimates.

The use of this equation rests on the assumption that the interest rate on which officials base planned, authorized, or actual expenditures is equal to the market rate currently paid by other cities floating bonds with the same characteristics as the bond issue which they contemplate. It is assumed that officials expect the current interest rate to prevail at the anticipated date of the bond sale.

Why is this equation used to estimate each city's borrowing costs? Why weren't the rates on comparable bond issues sold at the time officials made plans, authorizations, and actual expenditures used? The answer is that it was not until the hypotheses suggested by previous studies were tested that it was known which characteristics affect the cost of borrowing.

---

1. Roland I. Robinson, The Postwar Market for State and Local Government Securities, A Study by the National Bureau of Economic Research (Princeton University Press, Princeton, 1960) and articles by Frank E. Morris in the Investment Bankers Association's Statistical Bulletin.

2. The focus is on the response of expenditures planned and authorized for one year--1957. Therefore it is impossible to distinguish postponements from cutbacks.

Why weren't the interest rates paid by each city on the bonds it sold during the planning and authorization periods used to estimate the rates on which its officials based planned and authorized expenditures? There are two reasons. One is that cities adapt the types of bonds they issue to their needs so that they often do not plan to issue the same type of bond as they are currently selling. The other reason is that many cities do not sell bonds every year, and therefore in a random sample of cities some will not be selling a bond currently.

Another possible question is why the interest rate on the bond each city actually sold to finance authorized expenditures was not used to estimate the interest rate at the time expenditures were made. The reason is that the timing of the bond sale is affected by the volume and timing of expenditures which are themselves guided by fluctuations in the price of municipal bonds. Therefore the rate of interest on the actual bond issue reflects the response to credit conditions rather than the condition causing the response.

For all these reasons the testing of the hypotheses in Part 2 and the resulting equation for estimating the cost of municipal borrowing are both a precondition and an integral part of studying the interest elasticity of municipal capital expenditures.

Part 4 is devoted to a discussion of the conclusions that follow from testing the hypotheses in Parts 2 and 3.

In this study some areas are explored which Morris and Pickering did not chart. The principal innovation is the estimation of changes in capital expenditure plans which are made in response to tightened credit at different stages of the municipal decision process. Of secondary importance is the use of data on total capital expenditures, not just bond financed construction expenditures.

Some of the criticisms leveled at Morris and Pickering are also applicable to this study. Specifically, cutbacks and postponements are still indistinguishable. All that can be said is that a lower level of expenditures at a later stage of the decision process reflects cutbacks and/or postponements. However it is important to realize that the measures of planned and authorized expenditures used in this study permit the measurement of more cutbacks than the previous studies. Pickering used planned bond sales as a proxy for planned expenditures, but plans for bond sales--in the sense that the date of sale is set--are usually made subsequent to the planning and authorization of expenditures. Morris measured postponements and cutbacks from the cyclical pattern of actual expenditures. Who knows what the pattern of actual expenditures would have been had monetary conditions been different?

While an attempt is made to measure the substitutability between tax financed and bond financed expenditures, no attempt is made to measure the substitutability between construction expenditures and other types of capital outlays. However it does seem an improvement over previous research that reductions in all types of capital outlays are recorded.

Of course, the fact that all the data are taken from a period of tightening credit means that this study contributes no information about the impact of monetary policy on the shift of municipal capital expenditures to the period after the tightening of credit.

D. Nature of the Data

The hypotheses in Part 2 are tested with data on 2,000 bonds issued by states and other local governments besides municipalities.[1] This is done

1. Other local governments include counties, school districts, and special

to provide a large sample for studying the effects of investor risk on the cost of borrowing, and to ascertain whether other levels of government are treated by the market in the same way as municipalities. The final equation for estimating the cost of borrowing contains the answer to investors' preferences concerning the type of issuing agency and makes it possible for other researchers to estimate the rate of interest used by other levels of government in planning, authorizing, and executing capital expenditures.

Several private firms--Moody's and Standard and Poor's being most widely known--rate state and local borrowers according to their ability and willingness to repay loans. All state government and most local government bonds that are rated by Moody's enjoy one of the four top ratings (Aaa, Aa, A, and Baa); for the most part lower ratings are assigned to foreign government bonds. Bonds rated below Baa are excluded from the sample in Part 2 because there are so few of them. Unrated bonds are also excluded because the credit worthiness of the governments which issue them varies so widely.[1]

The data on planned, authorized, and actual capital expenditure in Part 3 were obtained through a mail survey of the 148 United States cities which keep capital budgets and whose bonds are rated Baa or better by Moody's. The survey was limited to cities whose bonds are so rated because the sample in Part 2 is restricted in that way. The survey was limited to cities which keep capital budgets because capital budget data make it possible to investigate some of the questions unanswered by previous studies.

Capital budget accounts are kept for all types of capital outlays--land and equipment as well as construction. Most cities keep separate accounts for bond financed and tax financed expenditures. Therefore capital budget data are ideally suited for an investigation of the impact of monetary policy on all types of capital outlays and of the incidence of monetary policy on tax financed expenditures relative to bond financed expenditures.

The capital budget data have the additional advantage of being objective. When asked for data, all officials respond from a uniform frame of reference and can date their decisions.

The limitations imposed on the survey population necessitate two assumptions if the sample inferences are to be applied to the population of all municipal borrowers. First it must be assumed that cities which keep capital budgets respond to changes in monetary conditions in the same way as cities which do not keep capital budgets. There is no evidence to controvert this assumption. However the kind of capital budget which shows long term plans over five or six years constitutes an advanced administrative technique. Cities in the survey which use such a budget may not therefore be representative.[2]

---

districts. Some authorities and commissions are administered by state officials and some by local officials.

1. Moody's does not assign ratings to bonds issued by governments that have less than $600,000 debt or fail to report all the information required to formulate a rating. Neither do they assign ratings to revenue bonds with no proven earnings, to bonds payable solely from special benefit assessments, or to bonds privately held as a whole. Moody's, Municipals and Governments.

2. The term capital budget is used in two senses. It is sometimes used to describe a long term capital improvement program; i.e., a five or six year projection of capital expenditures. It is also used to designate that part of the current budget which deals with capital outlays. Accounting for capital

Second it must be assumed that cities which issue bonds with no credit rating or a rating below Baa respond to tightened credit in the same way as cities which issue bonds with one of the four top ratings. The assumption that credit rating does not affect a city's response to monetary policy (except through the interest rate) is built into the regression model and therefore the exclusion of some credit ratings does not introduce another limitation. However the exclusion of cities that issue unrated bonds may bias the results on the side of less sensitivity to interest rate changes because municipalities which issue unrated bonds generally rely less on debt finance.

2. The Impact of Tightening of Credit on the Structure of State and Local Government Bond Interest Rates

The purpose of this part of the essay is to determine which types of bonds become relatively more expensive to issue when credit tightens. The bonds are typed according to seven characteristics which previous studies suggest define the structure of interest rates: maturity, optional feature, credit rating, size of issue, type of collateral, type of issuing agency and purpose of issue.

Because the structure of interest rates is multi-dimensional, in order to study the effect of one of the characteristics on the interest rate it is necessary to use a method of analysis which holds the values of the other characteristics constant. Some previous researchers dealt with this problem by selecting for study a sample of bonds for which at least some of the variables were constant.[1] In this study multiple regression analysis is used to separate out the effects of each of the characteristics on the interest rate and to analyze the impact of monetary policy on the structure of interest rates.[2] Before presenting the findings of the multiple regression analysis, I will discuss the foundations for the hypothesized relationships between the state-local bond rate and the seven structural characteristics and the foundations for the hypothesized impact of monetary policy on the structure of rates.

A. Hypotheses: The Structure of Interest Rates

(1) The interest rate increases with the maturity of the bond. The foundation of this hypothesis is that long term bonds carry greater risk of capital loss than short term bonds and that investors demand compensation for assuming the additional risk of holding long term bonds.

(2) A maturity date with certainty commands a lower interest rate than a maturity rate with uncertainty. The grounds for this differential are that a bond which the borrower can redeem before maturity--a callable or

---

outlays in the current budget is a standard operating procedure. Data taken from long term capital improvement programs are used to measure the impact of tightening credit on the planning of expenditures. Data taken from current budgets are used to measure postponements and cutbacks of authorized expenditures.

1. Op. cit., Roland I. Robinson and Frank E. Morris.

2. In a multiple regression, the relationships between a dependent variable and several independent variables are estimated. The independent variables are mathematically held constant so that the estimate of the relationship between the dependent variable and each of the independent variables is not subject to the error that would arise if each of the independent variables were used, as the sole explanatory variable.

optional bond--imposes on the investor the additional risk of having to liquidate his investment at a time when he can reinvest only at a lower yield. At the same time the call feature gives the borrower the opportunity to redeem a debt contracted at a high interest rate when interest rates fall and to refinance, thus making a saving in interest payments. The result is that the borrower is willing to pay for the privilege of an optional feature and the investor is willing to grant the privilege only at a price.

(3) The interest rate varies inversely with the credit rating assigned to the bond. High ratings are based on several measures of sound financial management, on evidence that the trend of government revenues is either rising or steady, and on evidence that short term fluctuations in government revenues are small.[1] It is alleged that investors believe there are substantial differences between the risks on Aaa, Aa, A and Baa bonds and that the interest rates on bonds in each of these categories reflect the risk differentials.

Frank E. Morris conducted a study of the interest rate differentials among Aaa, Aa, A, Baa, and unrated bonds.[2] He found that interest rates do vary inversely with credit rating and that the median of unrated bonds falls between the median of A and Baa bonds. The range of interest rates on unrated bonds is much wider than the range of rates on rated bonds.

Roland Robinson compared the term structures of Baa bonds and Aaa bonds and found that the slope of the Baa interest-maturity curve is steeper than the slope of the Aaa interest-maturity curve. He concluded that investors "compound" quality and maturity risks.[3] I am making no attempt to verify the "compound" hypothesis.

(4) The interest rate increases with the size of the bond issue. This hypothesis follows from the assumption that investors in state and local government bonds like to diversify the risks in their portfolios. Large bonds are usually issued by state and local governments who are frequent borrowers. The principal investors already have some bonds that are issued by frequent borrowers and ask compensation for the sacrifice in portfolio diversity which their purchase necessitates. Another consequence of the desire to diversify is that large issues must be placed with several rather than one or two investors. Dealers' handling costs rise accordingly and may be passed on in the form of higher interest charges to the borrower.

Frank Morris investigated the relationship between interest rate and size of issue and concluded that the size of issue does not influence the interest rate.[4] He found that large issues sell at lower interest rates than small issues, but that the interest differential can be explained by the fact that large issues as a group are of higher quality than small issues. The idea that the municipal bond market is equipped to handle all sizes of issues at the same cost is generally accepted.

(5) The interest rate on general obligation bonds is lower than the interest rate on revenue bonds. The reason for the differential, as in the

---

1. For a discussion of factors on which ratings are based, see David M. Ellinwood, "Guideposts to Success in Debt Management," Municipal Finance, XXIX, February 1957.

2. Frank E. Morris, "The Structure of Interest Rates on Municipal Bonds," IBA Statistical Bulletin, July 1958.

3. Op. cit., Robinson, pp. 185-187.

4. Op. cit., Morris.

case of the other structural characteristics, is a difference in risks. The legal power of recovering debts is greater for bonds secured by the full taxing power of the issuing government than for bonds secured by revenues from the sale of public services or from the rental of public buildings. The obligor of a general obligation bond is required to levy whatever taxes are necessary to defray his debt. The obligor of a revenue bond, however, has no legal obligation to repay the debt if project revenues prove insufficient. However, it is generally recognized that a government which defaults on a revenue bond casts doubt on its ability to manage finances soundly, so that a government would probably use general fund revenues to meet debt service requirements on a revenue bond before defaulting.

In his study of the market for revenue bonds, Roland Robinson found that the interest rate on revenue bonds is higher than the interest rate on general obligation bonds.[1]

In the sample of all state and local bonds issued in 1957 and 1958 Frank Morris found that general obligation bonds carry higher credit ratings than revenue bonds.[2] The fact that few revenue bonds carry high ratings makes it difficult to separate the effects of type of collateral from the effects of credit rating on the interest rate.

(6.1) The interest rate on bonds issued by state governments is lower than the interest rate on bonds issued by local governments. The foundation for this supposition is that states have a greater degree of financial autonomy than local governments; i.e., that local government authority may be exercised only with the sufferance of state authority. The differential represents a premium demanded by investors to insure against the risk that the state may initiate actions which affect the credit worthiness of a local government. An example of such an act by a state against a local government occurred in 1956 when the Virginia Assembly gave the Governor authority to withhold state aid from local governments which integrated their school systems.

(6.2) Investors prefer to invest in school district bonds rather than other local government bonds.[3] The argument behind this hypothesis is that investors think school district taxes would be paid even if other taxes had to become delinquent, because education strikes so close to the roots of all families. Yet in 1957 when New York school districts had difficulty in selling bonds, Comptroller Arthur Levitt seemed to suggest the hypothesis that school districts are discriminated against for some reason.[4] I shall attempt to resolve this question.

(7) The interest rate differs according to the purpose for which the bond sale proceeds are to be used. The justification for this hypothesis is similar to that used to support the idea of a differential rate of school districts; i.e., that investors prefer to invest in some kinds of projects rather than others because they believe if a municipality were short of funds it would pay off the debt for some projects before others. There is no reason

---

1. Op. cit., Robinson, p. 210.

2. Frank E. Morris, "The Pattern of Ratings on Municipal Bonds by Type of Issue and Size of Issue," IBA Statistical Bulletin, August 1959.

3. Winthrop S. Curvin, A Manual on Municipal Bonds, New York, Smith Barney, 1956.

4. "A Report of the Governor's Committee on the Marketing of School Bonds," Albany, New York, 1957.

to believe that educational purposes always take top priority. For example, the investment officer of a New Haven savings bank said that he bought only water bonds on the belief that a town could not survive without water and would therefore defray water debts first.

B. Hypotheses: The Impact of Monetary Policy on the Structure of Interest Rates

(8) The state-local interest rate (i.e., the average rate) changes more than one percentage point for every change of one percentage point in the interest rate on long term U. S. Government bonds. Figure 1, showing the fluctuations in the yields of Treasury and state-local bonds from 1951 through 1960, supports this hypothesis. After the Treasury-Federal Reserve Accord of 1951, the yields on state and local securities rose more than the yields on Treasury bonds. In the 1952-53, 1955-57 and 1959-60 tight money periods, yields rose more on state and local bonds than on Treasury bonds while the reverse was true in the 1953-54 and 1957-58 easy money periods.

Robinson lists three factors which he believes contribute to the sensitivity of state and local bond yields to changes in monetary conditions.[1] First is the narrow market for state and local bonds. The interest return on state and local bonds, unlike other bonds, is exempt from taxation by the Federal government, and in most cases is exempt from taxation by the state in which the issuing agency is located. The tax exemption feature makes state and local bonds especially attractive to investors who are subject to a high income tax (principally wealthy individuals, commercial banks, and fire, marine, and casualty insurance companies) or to investors who want to hedge against changes in the tax laws (life insurance companies and mutual savings banks). But the same feature drives down the pre-tax yield to such an extent that the bonds have much less appeal to investors who have little to gain from the tax exemption feature.

Second, none of the principal investors prefers state and local bonds to other investments. Individuals and fire, marine, and casualty insurance companies give top priority to equities. Commercial banks give preference to commercial loans and liquidity requirements (U. S. Government bonds). Life insurance companies and mutual savings banks list mortgages as their first investment choice. Only when investors have surplus funds or when the terms of trade between their first investment choice and state and local bonds become less favorable do they participate heavily in the state and local bond market.

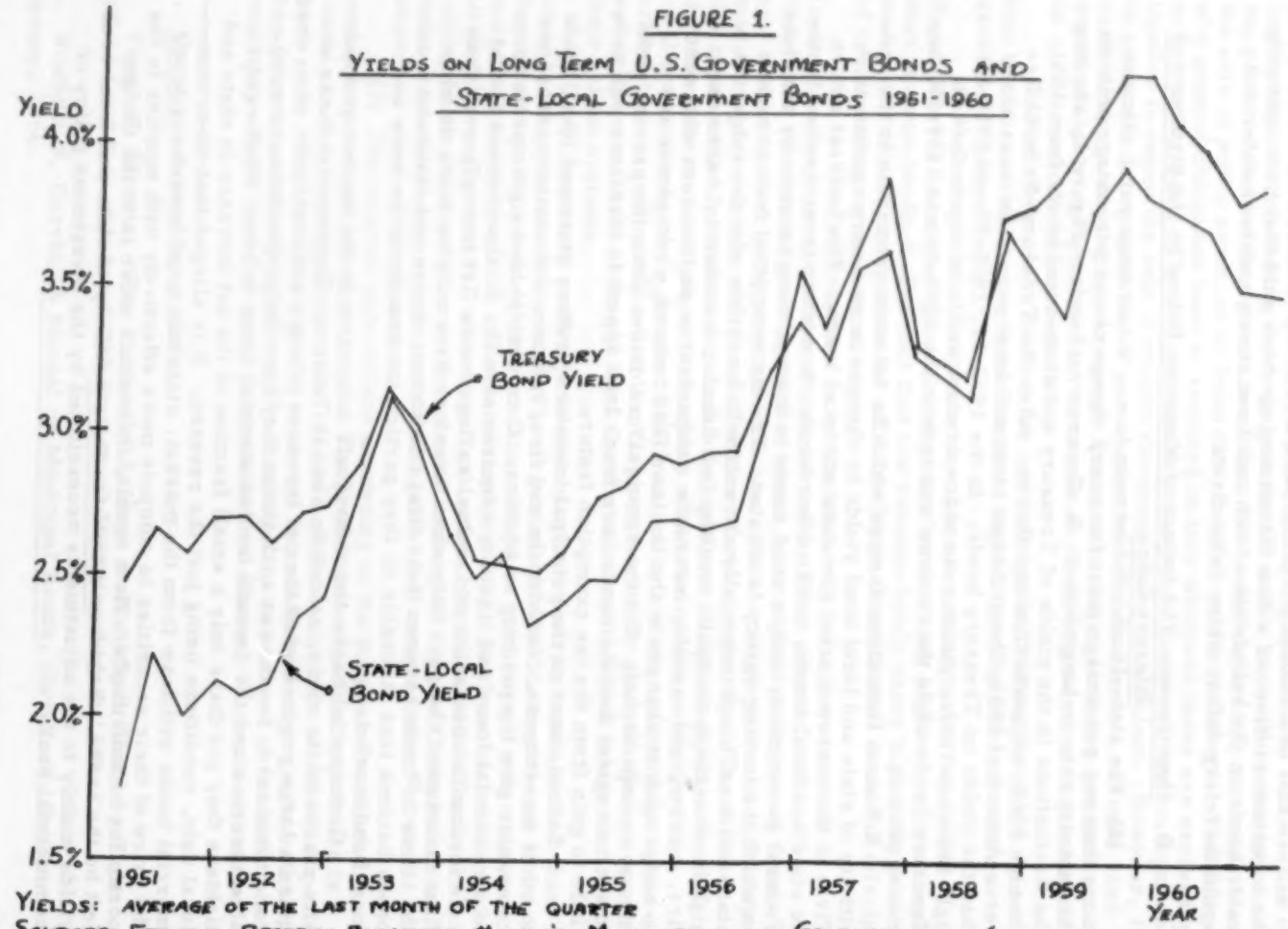
Commercial banks and individuals are active in the market at opposite phases of the cycle, as can be seen in Table 1. Commercial banks acquire a large proportion of the net increase in state and local debt when credit conditions ease, because at such times they have large quantities of available reserves and the demand for commercial loans is low. When credit tightens they purchase only a small fraction of the net increase in state and local debt, conditions being just the reverse. It is alleged that when commercial banks withdraw from the market, state and local borrowers lower the price of their securities to compete more effectively with equities in the portfolios of individuals. As a result, individuals move in to fill the gap left by the banks. Robinson attributes the sensitivity of state and local bond yields mainly to the adjustments necessitated by the investment policy of commercial banks.

---

1. Op. cit., Robinson, Chapter 3, "The Supply of Funds by Investors in State and Local Government Obligations."

FIGURE 1.

YIELDS ON LONG TERM U.S. GOVERNMENT BONDS AND STATE-LOCAL GOVERNMENT BONDS 1951-1960



YIELDS: AVERAGE OF THE LAST MONTH OF THE QUARTER

SOURCES: FEDERAL RESERVE BULLETIN; MOODY'S MUNICIPALS AND GOVERNMENTS (INDEX OF YIELDS ON GENERAL OBLIGATION BONDS)

TABLE 1

Net Acquisitions of State and Local Government Bonds

By Class of Purchaser

(billions of dollars and percent)

Year	1957				1958				1959				1960	
Quarter	1	2	3	4	1	2	3	4	1	2	3	4	1	2
Individuals**	$0.8	0.3	0.8	0.4	0.7	-0.2	0.6	0.3	0.5	0.7	1.2	*	0.5	0.3
	62%	30	62	36	37	-13	46	38	38	50	67	0	50	38
Commercial Banks	$0.2	0.2	0.2	0.4	0.6	1.3	0.5	0.2	0.2	0.2	.1	*	*	-0.2
	15%	20	15	36	32	81	38	25	15	14	6	0	0	-25
Non-bank Financial Institutions	$0.2	0.4	0.2	0.2	0.5	0.3	0.1	0.2	0.5	0.4	0.3	0.3	0.4	0.6
	15%	40	15	18	26	19	8	25	38	29	17	75	40	75
State and Local Governments	$0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1
	8%	10	8	9	5	13	8	13	8	7	11	25	10	13
Total	$1.3	1.0	1.3	1.1	1.9	1.6	1.3	0.8	1.3	1.4	1.8	0.4	1.0	0.8
	100%	100	100	100	100	100	100	100	100	100	100	100	100	100

* less than $50 million

** The net acquisitions of individuals are computed by subtracting the net acquisitions of the other classes of investors from the total.

Source: Table 4, "Sector Statements of Sources and Uses of Funds," Federal Reserve Bulletin, August 1959 and October 1960.

Third, the state and local bond market is particularly susceptible to inventory adjustments. When credit conditions tighten, underwriters must resell state and local issues quickly if they are to make a profit. Failing to turn over an issue quickly, an underwriter must either take the reduced profit (or loss) or carry the issue until credit conditions ease. Because underwriters find it difficult to place issues in such a market and costly to get credit to carry issues, they offer lower prices for new issues coming on the market and bid on fewer issues. The argument is reversed for easing credit conditions.

(9) Monetary policy has a differential impact on the structure of interest rates; specifically, the interest rates on high risk bonds rise (fall) more than the interest rates on low risk bonds when credit conditions tighten (ease). This hypothesis follows from the assumption that investors in state and local government bonds are diversifiers and from the fact that the seven hypothesized structural characteristics are measures of different kinds of risks. The argument is that when credit conditions tighten, an investor who diversifies his portfolio will use the opportunity for a higher return to reduce the riskiness of his portfolio and that when credit conditions ease, he will be willing to take on more risk to gain a higher rate of return. This kind of behavior would cause interest rates to move in the hypothesized fashion.

Do investors demand a greater absolute increase or a greater percentage increase in the return on high risk bonds when credit tightens? A greater absolute increase is a necessary (but not a sufficient) condition for a greater relative increase. The hypotheses are tested by measuring the increase in the interest rate in absolute percentage (or basis) points.[1] The relative increase is calculated from the regression equations used to test the hypotheses.

C. Statistical Formulation of Tests of the Hypotheses

In order to test the hypotheses about the structure of interest rates and the impact of monetary policy, two regression equations are estimated by fitting the equations to a sample of 2,000 state and local government bonds issued during the 1955-57 period of increasing credit restraint.

In equation (1) variables which represent the structural characteristics and monetary conditions are used additively to estimate the level of the state-local bond rate; i. e., it is assumed that each of the explanatory variables exerts an effect on the state-local bond rate which is independent of the effects exerted by other variables. Monetary policy is assumed to affect the interest rate on each type of bond in the same way.

$$(1)\quad N = b_k + b_m M + b_o O + b_r R + b_s S + b_c C + b_a A + b_p P + b_g G$$

N = "net interest cost," the interest rate on municipal bonds
b_k = constant term
M = maturity
O = optional feature: O = 1 callable
O = 2 non-callable (omitted category)
R = credit rating R = 1 Aaa
R = 2 Aa
R = 3 A
R = 4 Baa (omitted category)
S = size of issue

1. One percentage point is equal to 100 basis points.

C = collateral	C = 1 general obligation bond
	C = 2 revenue bond (omitted category)
A = issuing agency	A = 1 state government
	A = 2 county government
	A = 3 municipality
	A = 4 school district
	A = 5 special district
	A = 6 authorities and commissions (omitted category)
P = purpose	P = 1 education
	P = 2 roads and bridges
	P = 3 utilities
	P = 4 "other" (omitted category)

G = the long term U. S. Government bond rate, the index of monetary conditions

In equation (2) the variables which represent the structural characteristics are used both additively and multiplicatively to determine the level of the state-local bond rate. The cross products of the structural variables with the Government bond rate are added to the terms in equation (1). The introduction of these cross product terms permits testing the hypothesis that monetary policy has a differential impact on the structure of interest rates.

$$(2) \quad N = b_k + b_m M + b_o O + b_r R + b_s S + b_c C + b_a A + b_p P + b_g G + b_{gm} GM + b_{go} GO + b_{gr} GR + b_{gs} GS + b_{gc} GC + b_{ga} GA + b_{gp} GP$$

In Table 2, the nine null hypotheses and their alternative hypotheses are expressed symbolically in terms of the regression coefficients.

Hypotheses (1) through (6) and Hypothesis (8) are tested by means of the "t" test. The "t" test is a comparison of an estimated value of a regression coefficient with an hypothesized value (null hypothesis) of the same coefficient and a decision as to whether the discrepancy between the two is sufficiently large to reject the null hypothesis. The decision about whether the discrepancy is large is based on the probability of getting the sample value of the coefficient if the experiment of fitting the regression equation to successive samples were performed an infinite number of times if the null hypothesis were true. If the probability of getting the sample estimate is low, the null hypothesis is rejected.

The probability of drawing the sample estimate from the population of possible sample estimates is given by the sample "t" ratio--the ratio of the difference between the regression coefficient and the null hypothesis to the standard error of the regression coefficient--in relation to the probability distribution of "t" ratios. If the sample value of "t" is greater than or equal to a "critical value" of "t," the null hypothesis is rejected.

The selection of the critical value of "t" depends on the sample size and the nature of the hypothesis to be tested. In large sample studies such as this, the critical value of "t" is chosen so that 99% (or 95%) of experiments yield a value of "t" smaller than or equal to it. This means that only 1% (or 5%) of the time will the experimenter decide to reject the null hypothesis when in fact it is true. Such a mistake is called a Type I error.

If the alternative to the null hypothesis is that the difference between the estimated and null hypothesized value of the coefficient is not equal to zero, the region of the probability distribution for which the null hypothesis is to be rejected is split between the two tails of the distribution. The so-

called "two tail" test tests for positive and negative deviations of the observed from the null hypothesized value of the coefficient.

If the alternative hypothesis is that the difference between the estimated and the null hypothesized value of the coefficient is positive (or negative) a "one tail" test is needed. Only positive (or negative) deviations must be tested for. In the one tail test the region of the probability distribution for which the null hypothesis is to be rejected is concentrated in either the right or the left tail of the distribution.

Hypotheses (7) and (9) cannot be tested by means of the "t" test, because they require a simultaneous comparison of three or more coefficients. In these cases the "F" test is applicable. The "F" test jointly tests the deviations of one or more estimated parameters from a set of hypothesized parameters.[1] A high "F" ratio indicated that the null hypothesis should be rejected. In other words, it indicates that the null hypothesis is not supported by the data.

As in the case of the "t" ratio, the values of "F" for which the null hypothesis should be rejected are determined from a table of probabilities. The nature of the "F" distribution makes all "F" tests one tail tests (the two tails of the "t" distribution are concentrated in one tail of the "F" distribution).[2]

D. Results of the Tests

The results of the tests of the hypotheses are discussed below. Tables 3 and 4 give the estimated values of the coefficients, the standard errors, "t" and "F" for equation (1). Line 6 of Table 5 gives the "F" ratio for Hypothesis (9).

(1) The null hypothesis that the coefficient of the maturity variable is zero is rejected at the .01 level with a right tail test, confirming the alternative hypothesis that the state-local bond rate increases with the maturity of the bond.[3]

(2) The null hypothesis that the coefficient of the callable category is equal to the coefficient of the non-callable category is rejected at the .01 level with a right tail test, confirming the alternative hypothesis that callable bonds bear a higher rate of interest than non-callable bonds.

(3) The null hypothesis that the difference between the coefficients of the credit rating categories is zero is rejected at the .01 level with a left tail test. This confirms the alternative hypothesis that state and local governments with high credit ratings float bonds at a cheaper rate than governments with a low credit rating.

(4) The null hypothesis that the coefficient of the size of issue variable is zero is rejected at the .05 level with a right tail test, but is not rejected at the .01 level. The size of issue coefficient is so small that as far as most local governments are concerned, the popular notion that the market

---

1. The "F" ratio for one independent variable equals the square of the "t" ratio.

2. For a more detailed explanation of the "t" and "F" tests see Paul G. Hoel, Introduction to Mathematical Statistics, Second Ed., (John Wiley and Sons, New York, 1954).

3. To say that the null hypothesis is rejected at the .01 level is to say that the probability of a Type I error (rejecting the null hypothesis when it is true) is 1%.

TABLE 2

Hypotheses: Equations (1) and (2)

(1) H_0* : $b_m = 0$

H_1**: $b_m > 0$

(2) H_0 : $b_{o_1} - b_{o_2} = 0$

H_1 : $b_{o_1} - b_{o_2} > 0$

(3.1) H_0 : $b_{r_1} - b_{r_2} = 0$

H_1 : $b_{r_1} - b_{r_2} < 0$

(3.2) H_0 : $b_{r_2} - b_{r_3} = 0$

H_1 : $b_{r_2} - b_{r_3} < 0$

(3.3) H_0 : $b_{r_3} - b_{r_4} = 0$

H_1 : $b_{r_3} - b_{r_4} < 0$

(4) H_0 : $b_s = 0$

H_1 : $b_s > 0$

(5) H_0 : $b_{c_1} - b_{c_2} = 0$

H_1 : $b_{c_1} - b_{c_2} < 0$

(6.1) H_0 : $b_{a_1} - b_{a_i} = 0$

H_1 : $b_{a_1} - b_{a_i} < 0$

for i = 2, 3, 4, 5 and 6.

(6.2) H_0 : $b_{a_4} - b_{a_i} = 0$

H_1 : $b_{a_4} - b_{a_i} \neq 0$

for i = 2, 3, 5 and 6.

(7) H_0 : $b_{p_i} = 0$

H_1 : $b_{p_i} \neq 0$

for i = 1, 2, 3 and 4.

(8) H_0 : $b_g = 1$

H_1 : $b_g > 1$

(9) H_0 : $b_{gm} = b_{go} = b_{gr} = b_{gs} = b_{gc} = b_{ga} = b_{gP} = 0$

H_1 : Some coefficients of the cross product variables are not equal to zero.

* H_0 : The null hypothesis

** H_1 : The alternative hypothesis

---

does not charge a penalty for borrowing large amounts is valid. The penalty for large size becomes economically significant only when very large size issues (e.g., in the hundreds of millions) are involved.[1]

1. When size of issue is used as the sole explanatory variable, it does not

TABLE 3

Coefficients, Standard Errors, "t" and "F" Ratios: Equation (1)

	Variable	Coefficient	Standard Error	"t" or "F"
(M)	Maturity	+.027	.0015	17.45**
(O_1)	Callable#	+.071	.017	4.12**
(R)	Credit Rating			593.31**
(R_1)	Aaa	-.934	.029	-31.65**
(R_2)	Aa	-.708	.019	-37.67**
(R_3)	A	-.348	.017	-20.87**
(S)	Size of Issue	+.0016	.0008	2.05*
(C_1)	General Obligation	-.179	.023	-7.50**
(A)	Issuing Agency			6.93**
(A_1)	State	+.037	.042	.88
(A_2)	County	+.109	.038	2.85**
(A_3)	Municipality	+.127	.034	3.69**
(A_4)	School District	+.194	.038	5.13**
(A_5)	Special District	+.106	.044	2.39*
(P)	Purpose			2.13
(P_1)	Education	-.030	.021	-1.412
(P_2)	Roads	-.048	.020	-2.38*
(P_3)	Utilities	-.010	.019	- .53
(G)	Gov't. Bond Rate	+1.900	.023	84.43**

$S_u = .281$

$R^2 = .828$

** Significant at .01

* Significant at .05

"t" ratio measures difference between coefficient of dummy category and category omitted.

TABLE 4

Tests for Differences Between Regression Coefficients: Equation (1)

Difference Between Coefficients	Standard Error	"t"
$b_{r_1} - b_{r_2} = -.225$	.041	- 5.46**
$b_{r_2} - b_{r_3} = -.359$	.032	-11.12**
$b_{a_1} - b_{a_2} = -.072$	.072	- 1.00
$b_{a_1} - b_{a_3} = -.090$	.070	- 1.28
$b_{a_1} - b_{a_4} = -.157$	.070	- 2.22*
$b_{a_1} - b_{a_5} = -.069$	.075	- .91
$b_{a_4} - b_{a_2} = +.085$	.070	+ 1.20
$b_{a_4} - b_{a_3} = +.067$	.068	+ .99
$b_{a_4} - b_{a_5} = +.088$	.073	+ 1.21
$b_{gr} - 1 = +.900$	.023	+40.00**

* Significant at .05
** Significant at .01

(5) The null hypothesis that the coefficient of the general obligation bond category is equal to the coefficient of the revenue bond category is rejected at the .01 level with a left tail test. This confirms the alternative hypothesis that the cost of floating a general obligation bond is less than the cost of floating a revenue bond.

(6) The null hypothesis that the coefficient of the state government category is equal to the coefficient of the school district category is rejected at the .05 level, but is not rejected at the .01 level with a left tail test. In each of the other tests in which the coefficient of the state government cate-

explain a significant proportion of the variation in the state-local bond rate and the sign of the coefficient relating size of issue to the state-local bond rate is negative. A multiple regression including all "main effects" but the Aaa classification shows that it is the addition of the Aaa category which uncovers the small but significant positive effect of size of issue on the state-local bond rate.

gory is compared with the coefficient of a local government issuer, the null hypothesis is not rejected at the .01 level.

The hypothesis that the coefficient of the school district category is equal to the coefficient of each of the other local government categories is not rejected at the .01 level with a two tail test.

The sizes of the values of "t" in the tests suggest that the classification of issuing agencies may be improved by pooling states with authorities and commissions, by pooling municipalities with counties and special districts, and by leaving school districts in a separate class. In other words, with a new classification the null hypotheses may be rejected.[1]

(7) The null hypothesis that the coefficients of the purpose categories are equal to zero is not rejected at the .01 level with a two tail test. This means that the alternative hypothesis that the cost of borrowing does vary with the purpose for which the funds are intended is not confirmed.[2]

(8) The null hypothesis that the coefficient of the Government bond rate is equal to unity is rejected at the .01 level with a right tail test. This confirms the alternative hypothesis that the state-local bond rate is more sensitive to changes in credit conditions than the Treasury bond rate.

(9) The hypothesis that the coefficients of the cross product variables are equal to zero is rejected at the .05 level, but is not rejected at the .01 level. The addition of the cross product terms to the terms in equation (1) causes the standard errors of the regression coefficients in equation (2) to become very large in comparison with the standard errors in equation (1). The effect is to make it impossible to estimate the size of the change in the risk differentials accurately.

E. Analysis of Collinearity

The rise in the standard errors is due to collinearity between the independent variables.[3] Equation (2) is analyzed below to determine which of the cross product variables contribute significantly to the explanation of the

---

1. Many of the bonds which were classified as authorities and commissions were issued by state agencies. This may explain why the statistical test shows so little difference between the state category and the authorities and commissions category. It is interesting that a simple regression of the interest rate on the issuing agency variable shows that states pay a lower interest rate than authorities and commissions. This is probably because most bonds issued by states are general obligation bonds, while most bonds issued by authorities and commissions are revenue bonds.

2. In a regression of the interest rate on the purpose variable alone, the null hypothesis is rejected at the .01 level with a two tail test. This means that the purpose variable duplicates the explanatory effects of some of the other independent variables.

3. When two independent variables are collinear they move together in such a way that the part one plays in explaining the variation in the dependent variables cannot be distinguished from the part the other plays. In the case in point, the Government bond rate fluctuated over a relatively narrow range during the period covered by the regression sample, so that the values of the cross product terms differ only slightly from the values of the variables not multiplied by the Government bond rate. The values of some of the cross products also differ only slightly from each other, because they have the Government bond rate component in common.

variance in the state-local bond rate and which are in effect duplicating the explanation of one of the variables in equation (1) or of one of the other cross product variables. The results of the analysis are shown in Table 5.

TABLE 5

The Municipal Bond Rate: Analysis of Variance in Regression

	Regression	Sum of Squares	Degrees of Freedom	"F" Ratio
(1)	Overall Mean: Deviations	1000.805	2201	
(2)	Regression on M, O, R, S, C, A, P, G: Deviations	172.498	2185	
(3)	Regression on M, O, R, S, C, A, P, G: Effects (1) - (2)	828.307	16	655.75**
(4)	Regression on M, O, R, S, C, A, P, G, GM, GO, GR, GS, GC, GA, GP: Deviations	166.768	2170	
(5)	Regression on M, O, R, S, C, A, P, G, GM, GO, GR, GC, GA, GP: Effects	834.037	31	350.08**
(6)	Addition of GM, GO, GR, GS, GC, GA, GP: Effects (2) - (4)	5.730	15	1.76*
(7)	Regression on M, O, R, S, C, A, P, G, GM: Deviations	168.505	2184	
(8)	Addition of GM: Effects (2) - (7)	3.993	1	51.76**
(9)	Regression on M, O, R, S, C, A, P, G, GO: Deviations	170.895		
(10)	Addition of GO: Effects (2) - (9)	1.603	1	20.48**
(11)	Regression on M, O, R, S, C, A, P, G, GR: Deviations	170.554	2182	
(12)	Addition of GR: Effects (2) - (11)	1.944	3	8.29**
(13)	Regression on M, O, R, S, C, A, P, G, GS: Deviations	172.495	2184	
(14)	Addition of GS: Effects (2) - (13)	.003	1	.03
(15)	Regression on M, O, R, S, C, A, P, G, RC: Deviations	171.502	2184	
(16)	Addition of RC: Effects (2) - (15)	.996	1	12.68**

TABLE 5 (CONTINUED)

	Regression	Sum of Squares	Degrees of Freedom	"F" Ratio
(17)	Regression on M, O, R, S, C, A, P, G, GA: Deviations	171.685	2180	
(18)	Addition of GA: Effects (2) - (17)	.813	5	2.07
(19)	Regression on M, O, R, S, C, A, P, G, GP: Deviations	171.898	2182	
(20)	Addition of GP: Effects (2) - (19)	.599	3	2.54
(21)	Regression on M, O, R, S, C, A, P, G, GM, GR: Deviations	167.432	2181	
(22)	Addition of GM and GR: Effects (2) - (21)	5.066	4	16.04**
(23)	Regression on M, O, R, S, C, A, P, G, GM, GR, GC: Deviations	167.405	2180	
(24)	Addition of GC: Effects (21) - (23)	.027	1	.36
(25)	Regression on M, O, R, S, C, A, P, G, GM, GR, GO: Deviations	167.331	2180	
(26)	Addition of GO: Effects (21) - (25)	.101	1	1.32
(27)	Regression on M, O, S, C, A, P, G, GR: Deviations	170.706	2185	
(28)	Addition of R: Effects (27) - (11)	.152	3	.65

** Significant at .01
* Significant at .05

The first step of the analysis is to add each cross product variable separately to equation (1) to see if it reduces the unexplained variance in the state-local interest rate by a statistically significant amount.[1] Lines 7 through 20 of Table 5 show that the effect of adding the cross product involving either maturity, collateral, optional feature, or credit rating to equation (1) is to reduce the unexplained variance by a significant amount at the .01 level. The addition of the cross product involving each of the other variables in equation (1)--size of issue, type of issuing agency and purpose of

1. The effect of adding each cross product is estimated while all the other cross product variables are excluded from the equation.

issue--fails to reduce the unexplained variation significantly at the .01 level.

The standard errors of the "pure" maturity, collateral and call feature coefficients do not grow large when each of the corresponding cross product variables is added to equation (1), indicating that for these variables it is probably collinearity between the cross products which is responsible for the large standard errors in equation (2). However the standard errors of the pure credit rating coefficients do grow large when the credit rating cross products are added to equation (1), indicating that it is probably collinearity between the pure credit rating variables and the corresponding cross products which is responsible for the large standard errors in equation (2).

The second step of the analysis is to root out the remaining collinear relationships among the four cross product variables which passed the first stage tests. This is done by adding each cross product in succession to equation (1). Lines 21 and 22 show that the effect of adding the maturity cross product to a regression equation which includes the terms in equation (1) plus the credit rating cross products is to reduce the unexplained variance in the interest rate by a significant amount at the .01 level. Lines 23 and 24 show that the effect of adding the collateral cross product to the regression equation in line 21 is a negligible reduction in the unexplained variance. Addition of the optional feature cross product to the regression in line 21 also fails to reduce the unexplained variance (see lines 25 and 26). In the regressions in which three cross products are included (lines 23 and 25), the "t" ratios for the credit rating and maturity cross products are significant, but the "t" ratios for the collateral and the optional feature cross products are not significant.

The results of the second stage of the analysis of variance may be interpreted as showing that only the maturity and credit rating variables can be profitably added to equation (1). This does not mean that monetary policy does not have a differential impact on the interest rates within other "structural" categories. Rather it means that the differential impact on rates within the maturity and credit rating categories is larger than on the others and that the sample data to which the equations are fitted does not permit measuring smaller changes in the spread between high and low risk bonds. If the same hypotheses were tested on data drawn from a period in which there were wider fluctuations in the Government bond rate, it is possible that the differential changes in the interest rates on other types of bonds would be statistically significant.

The last step of the analysis is to follow up the suspected collinearity between the pure credit rating variables and the credit rating cross products. This is done by adding the pure credit rating variables to a regression which includes all the other variables in equation (1) plus the credit rating cross product variables (line 27) and testing the null hypothesis that the coefficients of the pure credit rating variables are equal to zero. The null hypothesis is rejected at the .01 level (see line 28), confirming that the credit rating cross product is sufficient to explain the effects of credit rating on the interest rate.

F. The Final Equation for Estimating State and Local Government Bond Interest Rates

The tests based on the coefficients estimated from equations (1) and (2) show that the estimate of the parameters of the state-local interest rate equation can be improved by omitting from equation (1) the purpose and credit rating variables, by regrouping the type issuer variable, and by adding the maturity and credit rating cross product variables. These changes

are shown in equation (3).

$$(3)\quad N = b_k + (b_m + b_{gm}G)M + b_o O + b_{gr}GR + b_s S + b_c C + b_a A + b_g G$$

where

A = 1 counties, municipalities, and special districts
A = 2 school districts
A = 3 states, authorities, and commissions (omitted category)

Table 6 restates the structural and monetary policy hypotheses in terms of the coefficients estimated from equation (3). Tables 7 and 8 give the estimated values of the coefficients, the standard errors, and "t." Table 9 gives the results of an analysis of variance in equation (3), showing that all the variables included in equation (3) contribute significantly to the explanation of the state-local interest rate. The results of these final tests of the hypotheses are discussed below.

(1) The null hypothesis that the coefficient of the maturity variable is zero is rejected at the .01 level with a right tail test, if the Government bond rate is evaluated at the sample mean: 3.04%. The sample values of the Government bond rate range from 2.76% to 3.73%. The null hypothesis cannot be rejected for values of the Government bond rate less than or equal to 2.30%.

By evaluating the coefficient of the maturity variable at some values of the Government bond rate above 2.30%, the hypothesized effects of maturity on the structure of the state-local bond rate can be seen. For example if a Government bond rate equal to 3% is assumed, a borrower must add approximately 2 basis points to total interest costs for each additional year of maturity; if a Government bond rate equal to 5% is assumed, the incremental maturity costs are equal to approximately 8 basis points.

The percentage change in interest costs for any given maturity can also be estimated from equation (3), if values are assumed for all the variables. For example, if the Government bond rate is assumed to increase from 3% to 5% (a percentage increase of 67%), and if values of a $2,000,000, non-callable, general obligation, A-rated bond issued by a municipality are substituted in the equation, the percentage increase in the interest rate on a bond with an average life of ten years is equal to 145%, twenty years is equal to 158%, and thirty years is equal to 163%. These equations are shown in Figure 2a. It is interesting to notice that whether the increase in the interest rate is measured in absolute or relative basis points, the proposition holds that it increases more on long term bonds than on short term bonds.

(2) The null hypothesis that the coefficient of the callable category is equal to the coefficient of the non-callable category is rejected at the .01 level, with a right tail test. The regression coefficient shows that borrowers pay 7 basis points more for the privilege of issuing a callable bond.

(3) The null hypothesis that the difference between the coefficients of the credit rating categories is equal to zero is rejected at the .01 level with a left tail test. If the Government bond rate is assumed equal to 3%, the spread between the interest rate on an Aaa bond and an Aa is 21 basis points, between an Aa bond and an A bond is 33 basis points, and between an A bond and a Baa bond is 33 basis points. If it is assumed that a change in monetary policy pushes the Government bond rate up to 5%, the spreads widen to 36, 55, and 56 basis points, respectively. This means that when the Government bond rate increases from 3% to 5%, the interest rate on an Aa bond, the interest rate on an Aaa bond rises by a smaller number of basis

TABLE 6

Hypotheses: Equation (3)

(1) H_0* : $b_m + b_{gm}\bar{G} = 0$

H_1** : $b_m + b_{gm}\bar{G} > 0$

(2) $H_0 : b_{o_1} - b_{o_2} = 0$

$H_1 : b_{o_1} - b_{o_2} > 0$

(3.1) $H_0 : b_{gr_1} - b_{gr_2} = 0$

$H_1 : b_{gr_1} - b_{gr_2} < 0$

(3.2) $H_0 : b_{gr_2} - b_{gr_3} = 0$

$H_1 : b_{gr_2} - b_{gr_3} < 0$

(3.3) $H_0 : b_{gr_3} - b_{gr_4} = 0$

$H_1 : b_{gr_3} - b_{gr_4} < 0$

(4) $H_0 : b_s = 0$

$H_1 : b_s > 0$

(5) $H_0 : b_{c_1} - b_{c_2} = 0$

$H_1 : b_{c_1} - b_{c_2} < 0$

(6.1) $H_0 : b_{a_3} - b_{a_i} = 0$

$H_1 : b_{a_3} - b_{a_i} < 0$

for i = 1 and 2

(6.2) $H_0 : b_{a_2} - b_{a_3} = 0$

$H_1 : b_{a_2} - b_{a_3} \neq 0$

(7) $H_0 : b_g + b_{gm}\bar{M} + b_{gr}\bar{R} = 1$

$H_1 : b_g + b_{gm}\bar{M} + b_{gr}\bar{R} > 1$

* H_0 : The null hypothesis

** H_1 : The alternative hypothesis

In testing Hypotheses (1) and (7) $\bar{G}$ and $\bar{M}$ are evaluated at the sample mean values. $\bar{R}$ is evaluated at R_3.

---

points than the interest rate on an A bond, and the interest rate on an A bond rises by a smaller number of basis points than the interest rate on a Baa bond.

The percentage change in the interest rate for each credit rating category can be estimated from equation (3), if values are assumed for all the variables. If the Government bond rate is assumed to increase from 3% to 5%, and if values of a $2,000,000, non-callable, general obligation, twenty year bond issued by a municipality are substituted in the equation, the percentage increase in the interest rate on a bond in the Aaa category is equal to 175%, in the Aa category is equal to 166%, in the A category is equal to 154%, and in the Baa category is equal to 145%. These equations are shown in Figure 2b. Contrary to the findings on the maturity risk differential, the proposition that the risk differential between high rated bonds and low rated bonds widens when credit conditions tighten holds if the differential is measured in

TABLE 7

Coefficients, Standard Errors, "t" Ratios: Equation (3)

Variable	Coefficient	Standard Error	"t"
M	-.069	.015	- 4.62*
GM	+.030	.005	+ 6.47**
O_1	+.068	.017	+ 4.04**
GR_1	-.294	.009	-32.57**
GR_2	-.222	.006	-38.83**
GR_3	-.111	.005	-21.70**
S	+.0016	.0007	+ 2.11*
C_1	-.178	.021	- 8.55**
A_1	+.106	.027	+ 5.93**
A_2	+.161	.023	+ 4.68**
G	+1.710	.056	+30.69**

$S_u = .277$

$R^2 = .8319$

** Significant at .01
* Significant at .05

absolute basis points, but not if the differential is measured in relative basis points.

(4) The null hypothesis that the coefficient of the size of issue variable is zero is rejected at the .05 level, but is not rejected at the .01 level with a right tail test. The regression coefficient shows that state and local governments pay an additional 16 hundredths of a basis point for every additional million dollars floated; i.e., a $100,000,000 bond bears an estimated interest cost 16 basis points higher than a $1,000,000.

(5) The null hypothesis that the coefficient of the general obligation bond category is equal to the coefficient of the revenue bond category is rejected at the .01 level with a left tail test. The differential between general obligation and revenue bonds, as estimated from equation (3), is 18 basis points.

(6) The null hypothesis that the coefficient of the states, authorities, and commissions category is equal to the coefficient of either of the local government categories is rejected at the .01 level with a left tail test. The hypothesis that the coefficient of the school district category is equal to the

TABLE 8

Tests for Differences Between Regression Coefficients: Equation (3)

Differences Between Coefficients	Standard Error	"t"
$b_m + b_{gm}\bar{G} = +.028$	.001	+19.59**
$b_{gr_1} - b_{gr_2} = -.072$	.013	- 5.50**
$b_{gr_2} - b_{gr_3} = -.111$	.009	-11.83**
$b_{a_1} - b_{a_2} = +.055$	.042	+ 1.31
$b_g + b_{gm}\bar{M} + b_{gr}\bar{R} = 1.945$	.022	+87.254**

** Significant at .01

TABLE 9

Municipal Bond Rate: Analysis of Variance in Regression

	Regression	Sum of Squares	Degrees of Freedom	"F" Ratio
(1)	Overall Mean: Deviations	1000.805	2201	
(2)	Regression on M, GM, O, GR, S, C, A, G: Deviations	168.216	2190	
(3)	Regression on M, GM, O, GR, S, C, A, G: Effects (1) - (2)	832.589	11	985.41**
(4)	Regression on O, GR, S, C, A, G: Deviations	198.971	2192	
(5)	Addition of M and GM: Effects (4) - (2)	30.755	2	200.20**
(6)	Regression on M, GM, O, S, C, A, G: Deviations	313.859	2193	
(7)	Addition of GR: Effects (6) - (2)	145.643	3	632.04**
(8)	Regressions on M, GM, O, GR, S, C, G: Deviations	170.916	2192	
(9)	Addition of A: Effects (8) - (2)	2.700	2	17.58**

** Significant at .01

FIGURE - 2A

THE DEPENDENCY OF THE MUNICIPAL BOND RATE ON MATURITY.

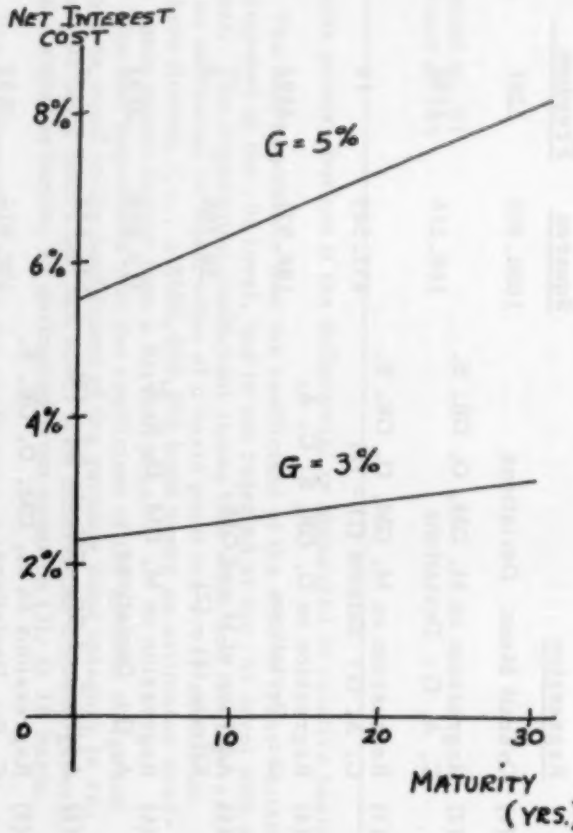
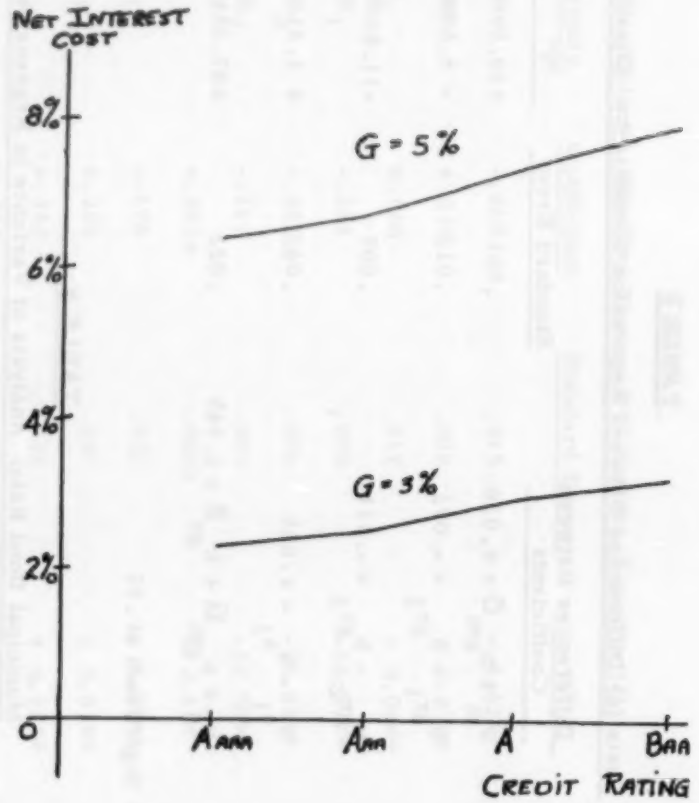


FIGURE 2B

THE DEPENDENCE OF THE MUNICIPAL BOND RATE ON CREDIT RATING.



coefficient of the counties, municipalities, and special districts category is not rejected at the .01 level with a two tail test. The regression coefficients show that investors demand a premium ranging from 11 to 16 basis points for holding any bond issued by a local government, as opposed to a bond issued by a state government, authority or commission.

(7) The null hypothesis that the coefficient of the Government bond rate is equal to unity is rejected at the .01 level with a right tail test. To see how much the state-local bond rate increases in response to an increase in the Government bond rate of 100 basis points, it is necessary to choose values for the maturity and credit rating variables in equation (3). For example, the interest rate on a twenty year A rated bond increases 220 basis points when the Government bond rate rises 100 basis points.

3. The Impact of Tightening Credit on Municipal Capital Expenditures

This part of the essay is addressed to the question: When credit tightens, do those municipalities which have planned or authorized the financing of capital expenditures by bonds which have become relatively more expensive to issue postpone and cut back a larger proportion of the total capital expenditures than other municipalities? The question is answered by relating the following variables in two groups of simple regressions:

(1) Dependent variable: the proportion of planned capital expenditures that is postponed and cut back.
Independent variable: the increase in borrowing costs experienced by cities between planning expenditures and revising plans.

(2) Dependent variable: the proportion of authorized capital expenditures that is postponed and cut back.
Independent variable: the increase in borrowing costs experienced by cities between authorizing and making expenditures.

There are six equations in each group. The dependent variable is calculated for bond financed, current tax financed and total capital expenditures. Alternative measures of the independent variable--the absolute and the relative increase in the municipal bond rate--are used with each dependent variable. The regression equations are used to test hypotheses and to give estimates of the impact of monetary policy on municipal capital expenditures.

A. Hypotheses

(1) When credit tightens the proportion of planned bond financed capital expenditures that is postponed and cut back is positively related to the increase in the cost of borrowing during the interval between planning expenditures and revising plans. This hypothesis implies that officials behave as if they make and revise plans on the assumption that the current interest rate will prevail at the time of the bond sale. It also implies that officials are unable or unwilling to raise taxes sufficiently to offset the additional costs of executing expenditures as planned.[1]

(2) When credit tightens the proportion of planned tax financed capital expenditures that is postponed and cut back is negatively related to the increase in the cost of borrowing during the interval between planning expenditures and revising plans. This hypothesis implies that when credit tightens officials increase the volume of tax financed expenditures relative to the volume initially planned. Because the method of finance is not legally tied to a particular project at this stage of the decision process, the increase in pay-as-you-go expenditures may reflect a change in the method of

1. These implications hold for Hypotheses (2) - (8) as well.

financing projects which were initially scheduled for bond finance.

(3) When credit tightens, the proportion of planned total (bond financed and tax financed) expenditures that is postponed and cut back is positively related to the increase in the cost of borrowing during the interval between planning expenditures and revising plans. This hypothesis implies that officials will find it necessary or preferable to reduce the size of the total improvement program. They will not increase the volume of tax financed expenditures by the amount of the decrease in the volume of bond financed expenditures.

(4) When credit tightens, the proportion of authorized bond financed expenditures that is postponed and cut back is positively related to the increase in the cost of borrowing during the interval between authorizing expenditures and actually making expenditures.

(5) When credit tightens, the proportion of authorized tax expenditures that is postponed and cut back is positively related to the increase in the cost of borrowing during the interval between authorizing expenditures and actually making expenditures. At this stage of the decision process the method of finance is legally tied to a particular project. Therefore the hypothesis implies that officials will prefer to postpone and cut back some tax financed projects in order to carry out more bond financed authorizations on schedule. They may borrow from the General Fund in order to reduce the size of the bond sale.

(6) When credit tightens, the proportion of authorized total expenditures that is postponed and cut back is positively related to the increase in the cost of borrowing during the interval between authorizing expenditures and actually making expenditures.

(7) Given a cross section of municipalities, the proportion of planned and the proportion of authorized expenditures that are postponed and cut back are positively related to the absolute increase in the interest rate. In other words, if all cities are faced with the same absolute increase in borrowing costs, officials in cities which initially expected to pay a high interest rate will postpone and cut back the same proportion of expenditures as cities which initially expected to pay a low interest rate.

(8) Given a cross section of municipalities, the proportion of planned and the proportion of authorized expenditures that are postponed and cut back are positively related to the relative increase in the interest rate. This means that given an across the board increase in absolute borrowing costs, cities which initially expected to pay a high interest rate will postpone and cut back expenditures less than those which initially expected to pay a low interest rate. Hypotheses (7) and (8) are alternatives.

B. Statistical Formulation of Tests of the Hypotheses

The hypotheses are subjected to confirmation or refutation by testing the null hypothesis that the slope coefficient (a_1) in equations (1) - (12) is equal to zero. The equations are related to the hypotheses as follows:

Hypothesis (1): equations (1) and (2)
Hypothesis (2): equations (3) and (4)
Hypothesis (3): equations (5) and (6)
Hypothesis (4): equations (7) and (8)
Hypothesis (5): equations (9) and (10)
Hypothesis (6): equations (11) and (12)
Hypothesis (7): equations (1), (3), (5), (7), (9), and (11).
Hypothesis (8): equations (2), (4), (6), (8), (10), and (12).

$$(1)\quad \left(\frac{P_{56} - P_{55}}{P_{55}}\right)_b = a_o + a_1 \left(N_{56} - N_{55}\right)$$

$$(2)\quad \left(\frac{P_{56} - P_{55}}{P_{55}}\right)_b = a_o + a_1 \left(\frac{N_{56} - N_{55}}{N_{55}}\right)$$

$$(3)\quad \left(\frac{P_{56} - P_{55}}{P_{55}}\right)_t = a_o + a_1 \left(N_{56} - N_{55}\right)$$

$$(4)\quad \left(\frac{P_{56} - P_{55}}{P_{55}}\right)_t = a_o + a_1 \left(\frac{N_{56} - N_{55}}{N_{55}}\right)$$

$$(5)\quad \left(\frac{P_{56} - P_{55}}{P_{55}}\right)_{b+t} = a_o + a_1 \left(N_{56} - N_{55}\right)$$

$$(6)\quad \left(\frac{P_{56} - P_{55}}{P_{55}}\right)_{b+t} = a_o + a_1 \left(\frac{N_{56} - N_{55}}{N_{55}}\right)$$

$$(7)\quad \left(\frac{A_{56} - E_{57}}{A_{56}}\right)_b = a_o + a_1 \left(N_{57} - N_{56}\right)$$

$$(8)\quad \left(\frac{A_{56} - E_{57}}{A_{56}}\right)_b = a_o + a_1 \left(\frac{N_{57} - N_{56}}{N_{56}}\right)$$

$$(9)\quad \left(\frac{A_{56} - E_{57}}{A_{56}}\right)_t = a_o + a_1 \left(N_{57} - N_{56}\right)$$

$$(10)\quad \left(\frac{A_{56} - E_{57}}{A_{56}}\right)_t = a_o + a_1 \left(\frac{N_{57} - N_{56}}{N_{56}}\right)$$

$$(11)\quad \left(\frac{A_{56} - E_{57}}{A_{56}}\right)_{b+t} = a_o + a_1 \left(N_{57} - N_{56}\right)$$

$$(12)\quad \left(\frac{A_{56} - E_{57}}{A_{56}}\right)_{b+t} = a_o + a_1 \left(\frac{N_{57} - N_{56}}{N_{56}}\right)$$

P = planned capital expenditures
A = authorized capital expenditures
E = actual capital expenditures
b = bond financed capital expenditures
t = current tax financed capital expenditures
N = "net interest cost," the municipal bond rate
55 = 1955

56 = 1956
57 = 1957[1]

Equations (1) - (6) are fitted to data on expenditure plans for fiscal 1957 which were made in 1955 and revised in 1956 and to estimates of the interest rates prevailing at the time plans were made and revised. Equations (7) - (12) are fitted to data on authorizations for 1957 which were made in 1956 and executed in 1957 and to estimates of the interest rates prevailing at the time expenditures were authorized and executed.

The independent variable in each equation reflects an increase in the cost of borrowing because 1955-57 was a period of increasing credit restraint.

The dependent variable is said to reflect postponements and cutbacks, because if tightening credit has an effect on the planning of total expenditures, then the volume of expenditures planned in 1956 for 1957 should be less than the volume of expenditures planned in 1955 for 1957, ceteris paribus. Similarly the difference between the volume of expenditures authorized for 1957 in 1956 and the volume actually expended in 1957 should represent postponements and cutbacks.

The dependent variable is formed by dividing the volume of postponements and cutbacks by the volume of expenditures initially planned or authorized. The sample cities vary widely in the scale of their expenditure programs. It is assumed that in response to a given change in the interest rate a city initially planning a large outlay will postpone a larger amount than will a city initially planning a smaller expenditure. On this assumption, expressing postponements as a proportion of planned expenditures reduces the deviations around the regression line.

It is assumed that the variables which determine the interest rate--such as those studied in Part 2--influence postponements only through the interest rate, e.g., there is no relation between credit rating and sensitivity to a given change in the interest rate.

One more question about the statistical formulation of the tests of the hypotheses remains to be answered. Is it necessary to fit the "planning" and "execution" equations to data on bond financed, tax financed, and total expenditures to find out how a city's response to tightening credit varies with the intended method of finance? Why is it not sufficient to look at either the two components of the total or the total and one component? Knowledge of the slope coefficients in two equations of a triplet is not sufficient to deduce the slope coefficient in the third equation because the relationship between changes in bond financed and tax financed expenditures in each sample city is not known.

C. Results of the Tests

The results of the tests of the hypotheses are discussed below. Table 10 shows the number of observations, the correlation coefficient, the standard error of estimate, and "t" for each regression. Table 11 shows which cities provided data for which regressions. The cities are identified by number rather than name because I promised to keep their identity anonymous.

(1) The null hypothesis as expressed in equation (1) is rejected at the .01 level with a left tail test. The null hypothesis as expressed in equation

---

1. The time subscripts date planning, authorization and actual expenditure decisions. Because cities make their decisions in different months, only the year of the decision is indicated in the notation. It should be noted that the value of N_{56} that appears in equations (7) - (12) is not equal to the value of N_{56} that appears in equations (1) - (6), because expenditures are authorized after plans are revised.

TABLE 10

Results of Tests: Simple Regressions

Equation					
Dependent Variable	Independent Variable	Sample Size	Correlation Coefficient	Standard Error of Estimate	"t" Ratio
(1) $\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_b$	$(N_{56}-N_{55})$	10	-.45	61.74	-1.432**
(2) $\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_b$	$\left(\frac{N_{56}-N_{55}}{N_{55}}\right)$	10	-.41	63.16	-1.267*
(3) $\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_t$	$(N_{56}-N_{55})$	10	.65	470.91	2.432***
(4) $\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_t$	$\left(\frac{N_{56}-N_{55}}{N_{55}}\right)$	10	.56	514.74	1.910**
(5) $\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_{b+t}$	$(N_{56}-N_{55})$	13	-.14	52.25	- .493
(6) $\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_{b+t}$	$\left(\frac{N_{56}-N_{55}}{N_{55}}\right)$	13	-.18	52.00	- .592
(7) $\left(\frac{A_{56}-E_{57}}{A_{56}}\right)_b$	$(N_{57}-N_{56})$	14	.09	29.29	.313
(8) $\left(\frac{A_{56}-E_{57}}{A_{56}}\right)_b$	$\left(\frac{N_{57}-N_{56}}{N_{56}}\right)$	14	.22	28.69	.779
(9) $\left(\frac{A_{56}-E_{57}}{A_{56}}\right)_t$	$(N_{57}-N_{56})$	14	.14	15.12	.493
(10) $\left(\frac{A_{56}-E_{57}}{A_{56}}\right)_t$	$\left(\frac{N_{57}-N_{56}}{N_{56}}\right)$	14	.24	14.81	.872
(11) $\left(\frac{A_{56}-E_{57}}{A_{56}}\right)_{b+t}$	$(N_{57}-N_{56})$	21	.09	21.55	.383

TABLE 10 (Continued)

Equation					
Dependent Variable	Independent Variable	Sample Size	Correlation Coefficient	Standard Error of Estimate	"t" Ratio
(12) $\left(\frac{A_{56}-E_{57}}{A_{56}}\right)$	$b+t\left(\frac{N_{57}-N_{56}}{N_{57}}\right)$	21	.25	20.94	1.137*

* Significant at .15
** Significant at .10
*** Significant at .05

(2) is rejected at the .15 level with a left tail test.[1] These findings confirm the alternative hypothesis that when credit tightens officials postpone and cut back the level of planned bond financed expenditures.

(2) The null hypothesis in equation (3) is rejected at the .05 level with a right tail test. The null hypothesis in equation (4) is rejected at the .10 level with a left tail test. These findings confirm the alternative hypothesis that when credit tightens officials increase the volume of tax financed expenditures relative to the volume initially planned.

(3) The null hypotheses in equations (5) and (6) are not rejected at the .15 level with a left tail test. These findings fail to substantiate the alternative hypothesis that when credit tightens officials postpone and cut back the level of total planned expenditures. An implication is that officials increase the volume of tax financed expenditures by enough to offset the decrease in the volume of bond financed expenditures. Table 11 shows such offsetting changes for cities (3), (4), (5), and (10).

(4) The null hypotheses in equations (7) and (8) are not rejected at the .15 level with a right tail test. These findings fail to substantiate the alternative hypothesis that when credit tightens officials postpone and cut back the volume of actual relative to authorized bond financed expenditures.

(5) The null hypotheses in equations (9) and (10) are not rejected at the .15 level with a right tail test. These findings fail to substantiate the alternative hypothesis that when credit tightens officials postpone and cut back the volume of actual relative to authorized tax financed expenditures.

(6) The null hypothesis in equation (11) is not rejected at the .15 level with a right tail test. The null hypothesis in equation (12) is rejected at the .15 level with a right tail test. That the null hypothesis in equation (11) is not rejected does not cast doubt on the inference that tightening credit has an impact on the execution of total authorized expenditures. Apparently it means that in this sample large absolute increases in the interest rate are correlated with initially high interest rates so that, while some cities cut back and postponed expenditures, there is little or no correlation between the magnitude of cutbacks and postponements and the absolute interest rate increase.

1. The tests are made on the basis of the probability of a Type I error equal to .15, because the sample size is small.

TABLE 11

Data To Which Equations (1) - (12) Are Fitted

City	$\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_b$	$\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_t$	$\left(\frac{P_{56}-P_{55}}{P_{55}}\right)_{b+t}$	$N_{56}-N_{55}$	$\frac{N_{56}-N_{55}}{N_{55}}$	$\left(\frac{A_{56}-E_{57}}{A_{56}}\right)_b$	$\left(\frac{A_{56}-E_{57}}{A_{56}}\right)_t$	$\left(\frac{A_{56}-E_{57}}{A_{56}}\right)_{b+t}$	$N_{57}-N_{56}$	$N_{57}-N_{56}$
Sample Size	10	10	13			14	14	21		
1	8.43%	-100.00	8.24	.46	16.59			13.68	.51	17.83
2	-10.19	- 28.67	-14.08	.33	12.08	32.52	24.33	32.22	.33	9.99
3	-15.54	172.40	31.52	.65	24.90			65.71	.32	9.94
4	- 2.86	3.06	0.00	.77	33.70					
5	-15.58	48.69	4.68	.47	16.66			21.80	.34	9.95
6	- 7.90	- 78.05	-22.38	.08	3.91	58.50	5.89	55.06	.41	20.76
7	122.03	0.00	122.03	.25	10.17					
8	- .39	- .39	- .39	.58	28.87	37.01	24.07	31.58	.29	11.09
9	169.65	- 24.46	114.84	.22	10.03	80.74	66.99	76.96	.46	19.29
10	- .95	1836.00	23.05	.93	35.03					
11			79.43	.29	12.86			36.67	.26	9.41
12			10.09	.13	6.45			37.27	.45	20.62
13			-40.16	.28	14.70			36.10	.46	19.36
14						58.74	24.56	55.28	.49	18.47
15						23.28	.00	22.90	.41	15.90
16						77.09	49.42	60.63	.48	19.10
17						18.99	20.28	19.14	.61	22.67
18						9.22	50.87	44.54	.48	19.18
19						11.81	44.84	30.33	.27	10.81
20						51.15	17.86	36.66	.28	10.29
21						82.52	50.36	74.21	.54	18.85
22						.05	9.99	6.11	.49	18.08
23						16.90	44.19	11.53	.52	15.52
24								1.41	.34	9.47

The sensitivity of total actual expenditures to monetary policy suggests that cutbacks and postponements of bond financed expenditures are reinforced by cutbacks and postponements of tax financed expenditures. Table 11 shows that those cities which reduced bond financed expenditures the most--cities (9), (16), and (21)--also reduced tax financed expenditures the most.

(7) and (8) In the tests of the hypotheses about the impact of tightening credit on the planning of expenditures it makes little difference whether the increase in the interest rate is measured in absolute or relative percentage points. Table 11 shows that those cities which experienced the largest absolute increase in borrowing costs between planning and revising plans also experienced the largest relative increase. This correlation explains why sensitivity to one of the measures is not appreciably greater than sensitivity to the other.

In the tests of the hypotheses about the impact of tightening credit on the execution of authorized total expenditures, the relative measure of the increase in the interest rate offers the better explanation of cutbacks and postponements. The explanation for this fact has been given under the discussion of the results of the tests of hypothesis (6).

D. Analysis of the Scatter of Observations Around the Regression Lines

Figures (3), (4), and (5) show the scatter of observations around the regression lines estimated by equations (2), (4), and (12).[1] Given that the hypotheses have been confirmed with a rather high probability of error, it is interesting to try to explain the scatter of observations with some collateral evidence obtained in the survey of municipalities.

The steep negative slope of equation (2) is attributable to two observations: cities (7) and (9). They revised plans markedly upward, while most of the other cities cut back or postponed the planned level of bond financed expenditures. Collateral evidence from these two cities indicates that the large increases in expenditures planned in 1956 relative to the volume planned in 1955 reflect small increases on a variety of projects.[2] There seem to be no unusual factors operating in either city to invalidate their inclusion in the sample. It may be inferred that their response followed from the fact that they were subjected to a relatively small increase in borrowing costs.

The steep positive slope of equation (4) is dominated by one observation: city (10). It is because the volume of pay-as-you-go construction in city (10) is negligible in the initial plans that the compensatory change appears as such a dramatic percentage increase. Given that three other cities which reduced bond financed expenditures made similar compensatory changes and that none of the three cities which increased bond financed expenditures also increased tax financed expenditures, there are grounds for believing the hypothesis that tax finance and bond finance are substitutes and that cities which experience the largest increases in borrowing costs increase reliance on tax finance. A letter from a survey city adds support to this hypothesis. The finance officer wrote, "Since the advent of the tight money period, the

1. The scatter of observations around equations (1) and (3) is similar to the scatter around equations (2) and (4).

2. City (9) sent me copies of the initial and revised capital improvement programs. I spoke with the finance officer in City (7) over the telephone.

FIGURE 3

$$(2): \left(\frac{P_{56}-P_{55}}{P_{55}}\right)_b = 72.06 - 2.47\left(\frac{N_{56}-N_{55}}{N_{55}}\right)$$

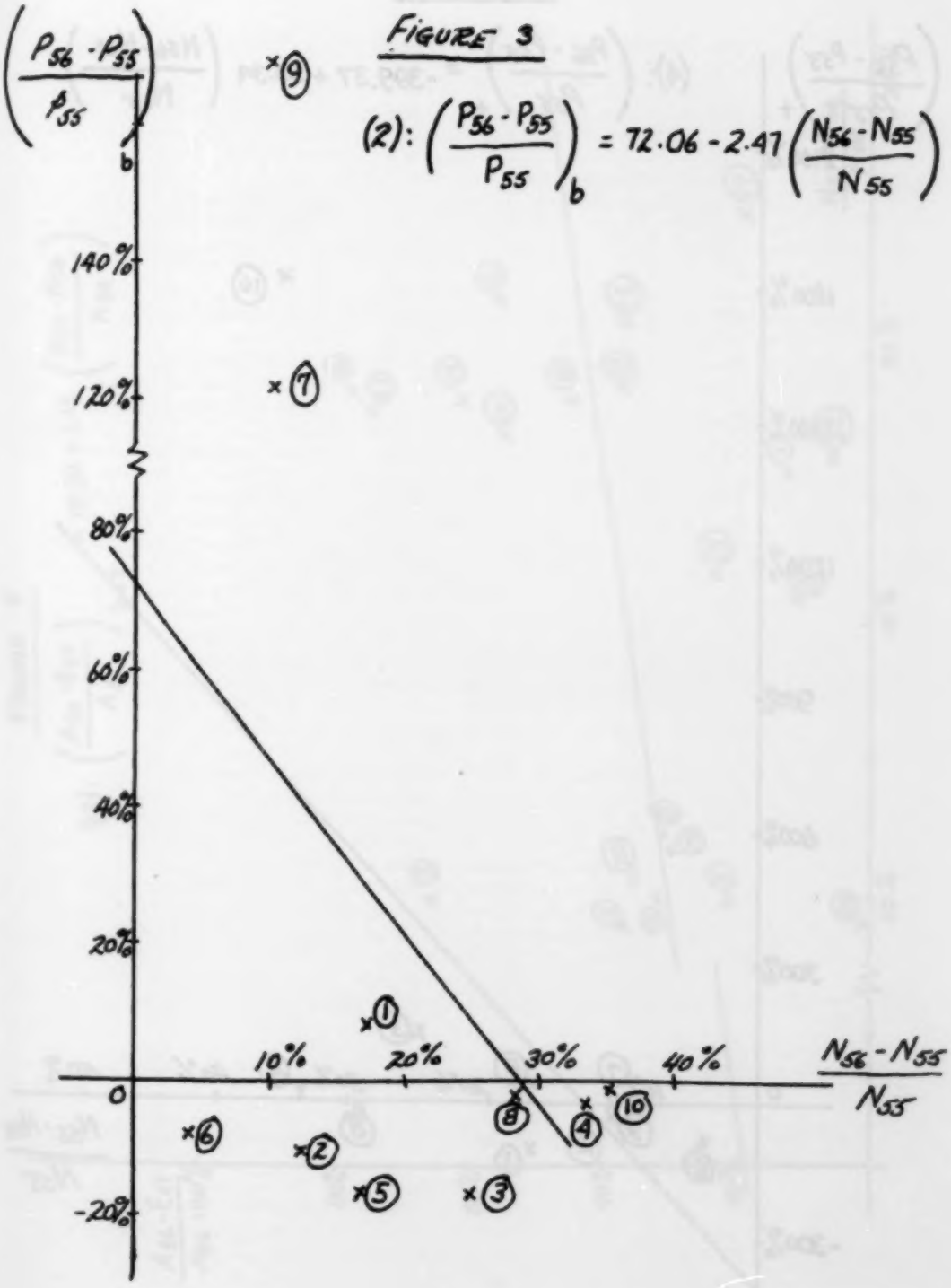


FIGURE 4

$$(4): \left(\frac{P_{56} - P_{55}}{P_{55}}\right)_+ = -399.37 + 30.34 \left(\frac{N_{56} - N_{55}}{N_{55}}\right)$$

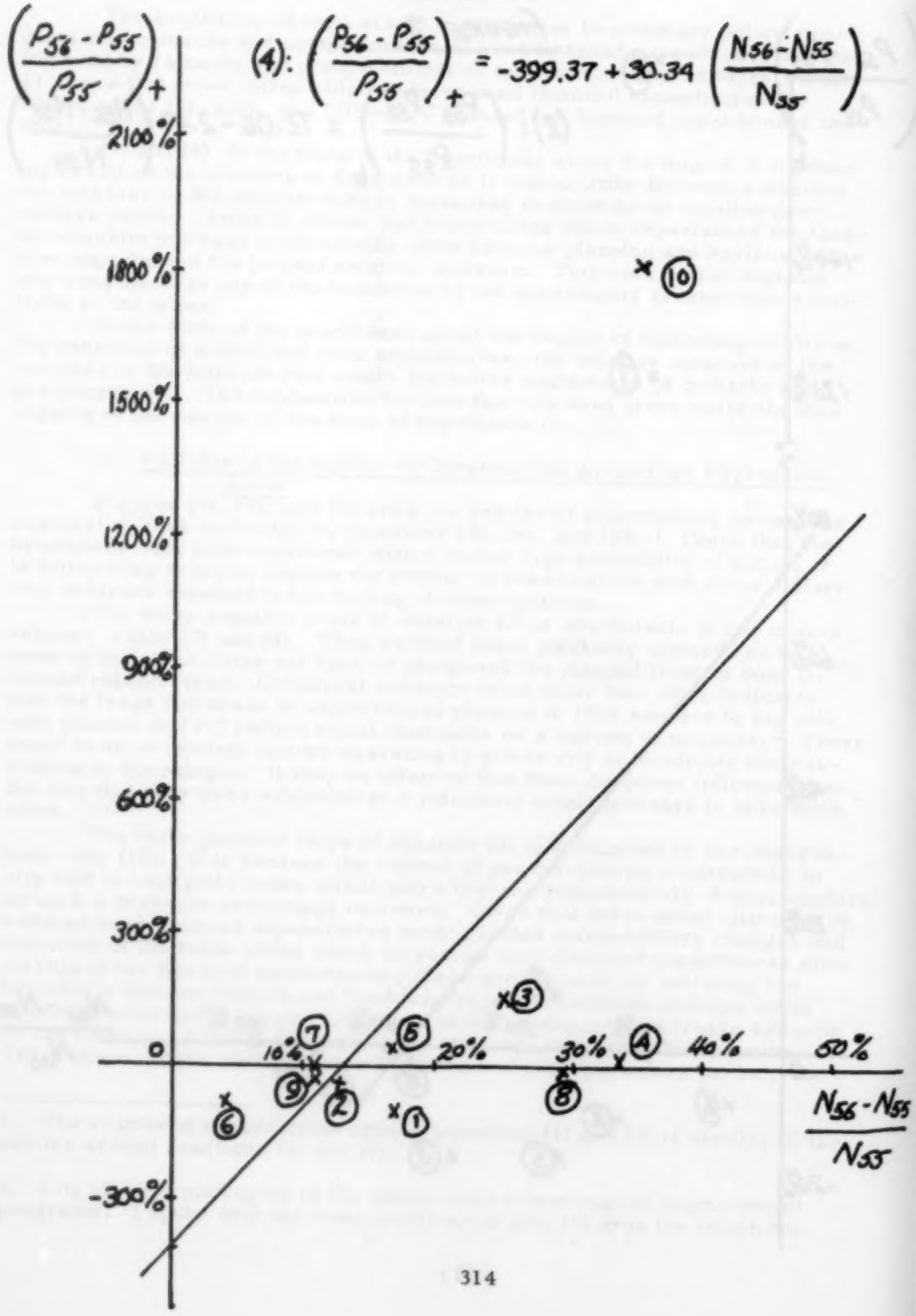
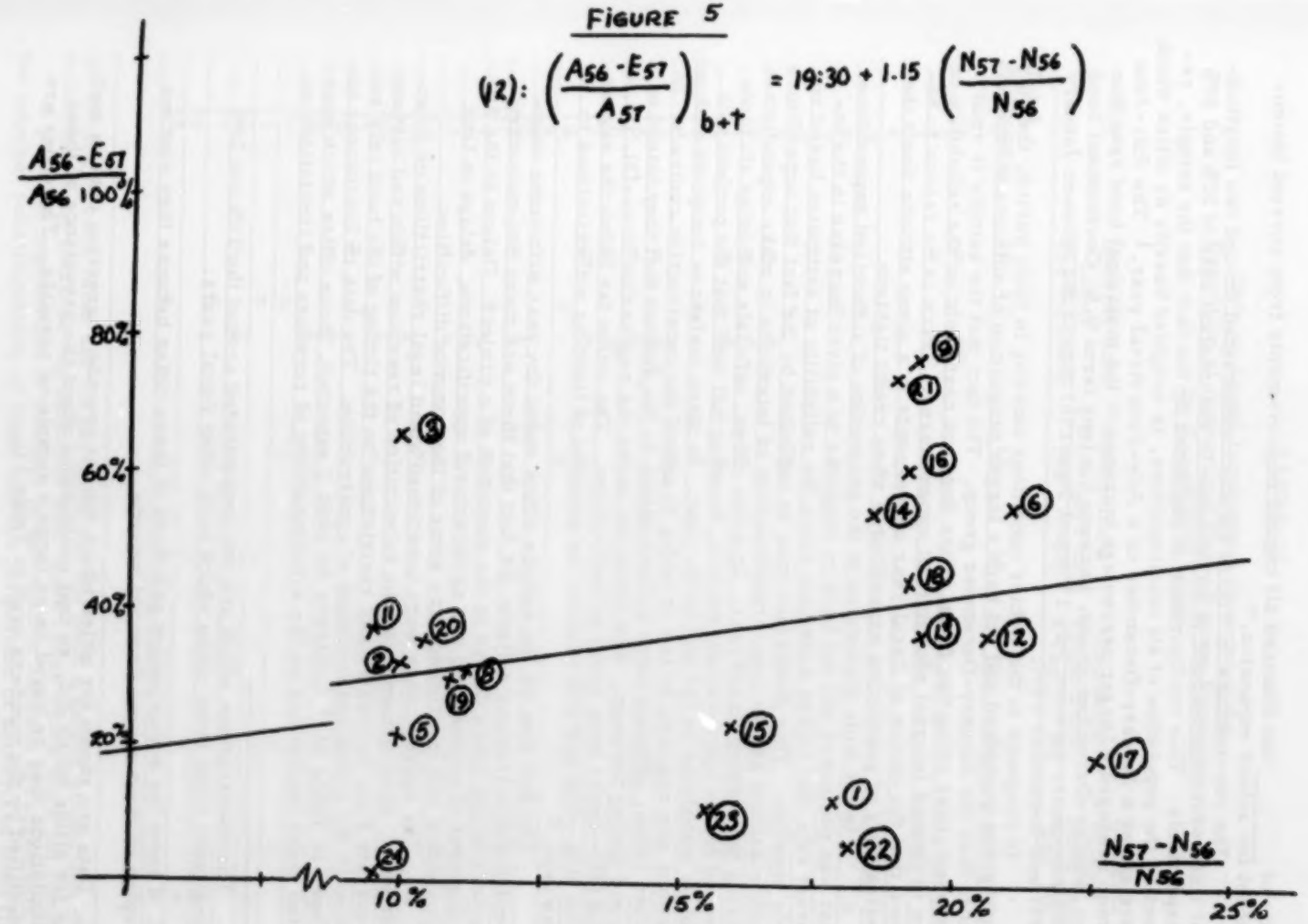


FIGURE 5

(12): $\left(\frac{A_{56}-E_{57}}{A_{57}}\right)_{b+t} = 19{:}30 + 1.15\left(\frac{N_{57}-N_{56}}{N_{56}}\right)$



City of _____ has financed all capital improvements from current income except for school expansion."

The observations in Figure (5) are concentrated around two longitudinal lines which correspond to an increase in the interest rate of 10% and 20% respectively. This configuration is explained by the fact that the sample, reflecting the population of all municipalities, is weighted heavily by cities which have either a January-December or a July-June fiscal year.[1] The July-June group averaged a larger percentage increase in the municipal bond rate than the January-December group, because the long term U.S. Government bond rate rose more between July 1956 and June 1957 than it did between January 1957 and December 1957.

In response to the tighter conditions existing in their period, the July-June group postponed and cut back a larger proportion of authorized expenditures than the January-December group. The fact that the sample is split between cities using two fiscal years does not cast doubt on the reliability of the estimated interest elasticity of expenditures. There is no reason to believe that the choice of fiscal year corresponds to a given attitude about the desirability of postponing expenditures when credit tightens.

It is the wide dispersion in the proportion of authorized expenditures that was postponed and cut back in response to a given increase in the interest rate that raises questions about the reliability of estimates based on this equation. This dispersion may be explained by the fact that expenditure authorizations are an inaccurate measure of intentions to make expenditures during the coming fiscal year. In some cities, officials authorize all funds needed for an improvement project, knowing full well that the project will not be completed by the end of the year. In many instances complete authorizations are required by law in order to award the construction contract. In other cities, officials limit authorizations to the amount that they intend to spend during the year. The cities far above the regression line--(3), (9), and (21)--make open ended authorizations. The cities far below the regression line--(22) and (24)--follow the practice of limiting authorizations to one year.[2]

The fact that all the sample cities ended the year with some unexpended authorizations reflects the fact that there are many non-monetary factors which cause delays in the execution of a project. Delays in the development of engineering and architectural specifications, delays in land acquisition and in litigation are some of the common difficulties.

The survey cities were questioned about legal restrictions on interest rates, arrangements for the substitution of revenues within and between separate Fund Accounts, and restrictions on the timing of the bond sale relative to the scheduled initiation of construction. The data on limitations are shown in Table 12.[3] Contrary to what I expected, those cities which operated under limitations on the substitutability of revenues and limitations on

---

1. The observations which are not concentrated around the 10% and 20% longitudes stem from cities which have other fiscal years.

2. I know the authorization practices of these cities because they sent me copies of their annual financial reports.

3. Data are shown for cities which did not provide regression data as well as for cities which did, so that conclusions about the prevalence of these limitations may be based on as large a sample as possible. The cities are identified by the numbers used in Table 11.

TABLE 12

Data on Legal Limitations Governing the Execution of Authorized Expenditures Provided by Respondent Cities

City	Interest Rate Limit			Substitution Between Funds		Consolidated Bond Fund		Contract Without Bond Sale	
	Yes	No	Limit	Yes	No	Yes	No	Yes	No
Total Responses	18	12		16	14	8	22	15	14
1	x		5%		x		x	x	
2	x		6	x		x		x	
3									
4									
5		x			x		x		x
6		x		x		x		x	
7									
8		x		x		x		x	
9	x		6		x		x	x	
10	x		5	x		x		x	
11	x		6	x			x		x
12		x		x		x		x	
13									
14	x		4.5		x		x		x
15	x		6	x			x		x
16	x		5		x		x		x
17	x		6	x		x		x	
18		x		x			x		x
19									
20	x		6		x		x	x	
21		x		x			x	x	
22	x		6	x			x	x	
23	x		6		x		x		
24		x		x			x		x
25	x		5		x	x		x	
26	x		6	x		x		x	
27	x			x			x		x
28	x		6		x		x		x
29	x		6	x			x		x
30		x		x			x	x	
31		x			x		x		x
32	x		6		x		x	x	
33		x			x		x		x
34					x		x		x
35		x			x		x		x

the timing of bond sales did not postpone and cut back a larger proportion of authorized bond financed expenditures than those cities which were not limited.

A possible explanation of the lack of correlation between arrangements for the internal substitutability of funds and the proportion of authorized bond fi-

nanced expenditures that was postponed is that those cities which have limited substitutability of internal funds find it easy to obtain short term credit externally. It may also be that short term borrowing plays a larger role than internal finance in helping officials to carry out expenditures as planned. Several cities volunteered the information that they relied on short term borrowing to tide them over the period of peak interest rates, while only one mentioned reliance on the city's own cash position.

D. An Estimate of the Impact of Tightening Credit on the Execution of Total Authorized Capital Expenditures

Given that equation (12) is based on a small sample obtained from a selected population within the broad population of municipalities in the United States and given that the standard error of estimate is rather high, any estimate of the aggregate response of municipalities to tightening credit should be viewed with caution. The reader having been forewarned, some estimates of the impact of tightening credit in 1957 and 1959 are presented for consideration.

Moody's index of yields on general obligation bonds is taken as a rough approximation of the average interest rate on which municipal officials base authorized and actual expenditures.[1] It is assumed that officials based authorized expenditures for calendar 1957 (59) on the average rate during the last quarter of 1956 (58). It is also assumed that they continually adjusted actual expenditures to changes in the interest rate so that the average interest rate prevailing at the time expenditures were actually made is represented by the average rate during 1957 (59).[2] Using these definitions the estimates of N are as follows:

$$N_{56} = 3.36\% \; (N_{56}\text{-}N_{57})/\, N_{56} = 5.95\%$$

$$N_{57} = 3.56\%$$

$$N_{58} = 3.62\% \; (N_{58}\text{-}N_{59})/\, N_{58} = 3.31\%$$

$$N_{59} = 3.74\%$$

All municipalities in the United States actually spend $3,653 million on total capital outlays in the fiscal year ending in 1957. In 1959 they spent $3,253 million.[3] If a rather bold simplifying assumption is made that all municipalities have a fiscal year that coincides with the calendar year, this information can be used in equation (12) to solve for authorized expenditures and the volume of expenditures that were cut back and postponed.[4] The re-

1. Equation (3) in Part 2 is not used to estimate the N's, because it is applicable only to individual borrowers.

2. These assumptions were also used in the estimation of the regression coefficients in Part 3.

3. Compendium of City Government Finances in 1959, U.S. Department of Commerce, Bureau of the Census.

4. The fiscal year followed by municipalities is not standardized. If the survey sample is representative, probably around half have a January-December fiscal year; and the other half a July-June fiscal year.

sulting solutions are:

A_{56} = \$4,404 million $A_{56}-E_{57}$ = \$1,151 million

A_{57} = \$4,750 million $A_{58}-E_{59}$ = \$1,097 million

Notice that not all these postponements and cutbacks are attributable to tightening credit. Had there been no change in credit conditions over the periods considered, officials would still have postponed and cut back about 19% of total authorized expenditures.[1] In 1957 the 6% increase in borrowing costs prompted officials to postpone and cut back an additional 7% of expenditures. In 1959 the 3% increase in the municipal bond rate caused officials to cut back expenditures by an additional 4%. This means that the volume of postponements and cutbacks due to tightening credit was about \$302 million in 1957 and \$181 million in 1959.

4. Conclusions

By way of conclusion I shall answer the two questions posed in Part 1: (1) When credit tightens, which types of bonds become relatively more expensive to issue? (2) Do those municipalities which have planned or authorized the financing of expenditures by bonds which have become relatively more expensive to issue cut back and postpone a larger proportion of total capital expenditures than other municipalities?

The first general hypothesis in Part 2--that the structure of state-local interest rates results from investors balancing different types of risk against the rate of return--is confirmed by the sample of state and local government bonds sold during the 1955-57 period of increasing credit restraint. The coefficients estimated by equation (3) in Part 2 indicate that low risk bonds bear a lower rate of interest than high risk bonds. They show that interest rates on short term bonds are lower than rates on long term bonds for all values of the long term U.S. Government bond rate greater than or equal to 2.30%; that interest rates on non-callable bonds are lower than rates on callable bonds; that interest rates on bonds with high credit ratings are lower than interest rates on bonds with low credit ratings; that interest rates on small bonds are lower than rates on larger bonds; that interest rates on general obligation bonds are lower than rates on revenue bonds; and that interest rates on state bonds are lower than rates on local government bonds.

The data do not support the hypothesis that the cost of borrowing varies according to the purpose for which the funds are intended; i.e., that investors consider bonds issued for certain purposes especially risky. They also show no difference in the interest costs borne by school districts as compared with the costs borne by municipalities and other local governments.

The second general hypothesis--that monetary policy has a differential impact on the structure of interest rates--is also confirmed. However in this sample it is possible to measure the differential impact within only two of the six risk categories--maturity and credit rating-- without running into problems of multicollinearity. The hypothesis that a given increase in the Government bond rate causes a larger increase in the interest rate on long term state-local bonds than on short term bonds is confirmed, whether the increase in measured in absolute or relative basis points. The hypothesis that a given increase in the Government bond rate results in a larger increase in the cost of issuing low rated bonds than in the cost if issuing high rated

1. Non-monetary causes of postponements are listed on page 316.

bonds is confirmed only if the increase is measured in absolute basis points. The percentage increase in the cost of issuing low rated bonds is less than the percentage increase in the cost of issuing high rated bonds.

The results of the tests of these hypotheses provide the information for determining which types of bonds become relatively more expensive to issue when credit tightens. If the increase in borrowing costs is measured in absolute basis points, it is long term low rated bonds which become relatively more expensive to issue. If the increase is measured in relative basis points, it is bonds which have less risky characteristics in all categories except maturity that become relatively more expensive. A given absolute increase in borrowing costs results in a larger percentage increase in the cost of issuing bonds with initially low interest rates than bonds with initially high rates.

The results of the tests of the hypotheses in Part 3 suggest that, among municipalities which have planned or authorized the financing of some capital expenditures by long term borrowing, an increase in the municipal bond rate affects the planning and execution of capital expenditures differently.

At the planning stage of the decision process, a tightening of credit has a negligible impact on total capital expenditures. The regressions of planned current tax financed expenditures and planned bond financed expenditures on the increase in the municipal bond rate indicate that municipalities cushion the impact of tightening credit by increasing reliance on tax finance and decreasing reliance on bond finance.

The hypothesis is confirmed that, at the execution stage of the decision process, municipal officials postpone and cut back the total volume of actual relative to authorized capital expenditures to a greater extent the greater is the relative increase in their borrowing costs. However when credit tightens the proportion of authorized bond financed expenditures and the proportion of authorized tax financed expenditures that are postponed and cut back are each statistically insignificant when measured separately. My interpretation of these findings is that officials distribute the burden of adjustment to tighter conditions between tax financed and bond financed expenditures, achieving the objective of staying within the budget constraint while making the adjustment in each expenditure category relatively easy to bear.

The hypothesis that tightening credit causes postponements and cutbacks of actual relative to authorized total expenditures is not confirmed when the increase in borrowing costs is measured in absolute terms instead of relative terms. However this does not cast doubt on the inference that tightening credit has an impact on the execution of total authorized expenditures. Apparently it means that in this sample large absolute increases in the interest rate are correlated with initially high interest rates so that, while some cities cut back and postponed expenditures, there is little or no correlation between the magnitude of cutbacks and postponements and the absolute interest rate increase.

In the tests of the hypotheses about the impact of tightening credit on the planning of expenditures, it makes little difference whether the increase in the interest rate is measured in absolute or relative terms. This is because in this sample those cities which experienced the largest absolute increase in borrowing costs between planning and revising plans also experienced the largest relative increase.

The question of the relative versus the absolute increase in the interest rate as alternative explanations of the proportion of expenditures that

is cut back and postponed is important in analyzing the incidence of tightening credit on different kinds of state and local borrowers. For example, if the absolute increase offered the better explanation of the observed response to tightening credit, it could be inferred that, other things being equal, cities with low credit ratings are "hit harder" by tightening credit than cities with high credit ratings. If the relative increase offered the better explanation, the opposite inference would be justifiable. In view of the fact that the absolute increase offers the better explanation of the variation of some dependent variables and that the relative increase offers the better explanation of the variation of others, further research on the "absolute-relative" question would be desirable.

Summary: When a world of certainty is assumed, the application of profit maximization theory to commercial banks proves sterile. Useful analysis of bank portfolios must recognize the importance of uncertainty to bank operations--uncertainty about future deposit levels, about market values of securities, about loan defaults, and about the bank's ability to achieve the portfolio it desires. Once these elements of uncertainty are explicitly considered, the assumption of expected-profit maximization can aid our understanding of bank behavior.

In this model, the bank is assumed to hold a fixed fraction of its deposits in cash-like forms, and to divide the remainder of its asset portfolio between securities and loans. Securities sales are preferred as a means of meeting any net withdrawals of deposits that occur, but securities earn less than loans and are subject to possible capital value change. An optimum division between these two assets is found, and from it is derived the marginal (opportunity) cost of making loans. In effect, this latter is the bank's supply curve of loans. An analysis is made of the impact upon the loan supply curve of changes in securities earning rates, in possible fluctuations in deposits or securities prices, and in the parameters of central bank policies.

The model is basically an effort to strengthen the micro-economic foundations upon which all aggregative monetary theory must be constructed.

This essay was prepared in 1960 from a dissertation submitted in 1957 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Research Advisor at the Institute of Development Economics, Karachi, Pakistan.

Dissertation Committee: Professors James Tobin (Chairman), Thomas C. Schelling, Jacob Marschak, Raymond P. Powell.

*

The author wishes to thank James Tobin for continued guidance and help with this paper. Useful suggestions have also been given by Donald Hester, Arthur Okun, Raymond Powell, Herbert Scarf, T. N. Srinivasan, and Robert Summers.

R. C. P.

A MODEL OF BANK PORTFOLIO SELECTION

Richard C. Porter

1. Introduction

Over the course of the last century, the implications of the assumption of profit maximization for the behavior of the firm have been tracked down in ever greater detail. Curiously, however, this firm has almost always been a seller of non-financial goods; banking has been studiously exempted from the application of such theory. The exemption is curious because the commercial bank seems in many respects more likely to fit the conditions of such static theory than the product manufacturer. The "method" of production and the "product" itself do not change, and hence the unpleasant necessity of neglecting some of the most interesting features of markets in order to devise marginal conditions does not arise. Even the proverbial conservatism of bankers is a prop to such theory, for it may well make banking less prone to upsetting expectational factors than other markets.

The reason for this neglect of banking probably lies in the implication of straightforward profit maximization: that the bank should acquire a portfolio consisting entirely of the asset whose yield (less any costs of maintainence or acquisition) is greatest.[1] But this procedure misses the very essence of banking, which is to "borrow short and lend long." Thus, the "profit" which a bank derives from its portfolio must be interpreted in terms of not only the money return but also the liquidity and capital certainty which the portfolio offers. There is no reason why the concepts of profit maximization cannot be applied to bank operations, provided that "profit" is conceived in this broader sense.

The crux of bank operations is uncertainty, and hence any reference to profits must be in a probabilistic sense. In this paper, it will be assumed that the bank considers the expected value of its profits (i.e., additions to surplus during the planning period) under various conditions of risk,[2] principally that of change in size of deposits. The problem is somewhat analogous to recent demand-risk-inventory theory for the selling firm,[3] where cash (and other assets readily convertible into cash) represents inventories, the carrying cost of these "inventories" is the surrender of earning power, and various penalties are incurred for insufficient "inventories."

This approach to commercial bank operations is not new, having been first indicated by Edgeworth in 1888,[4] although at that time bankers still considered loans to be liquid (in the sense of self-liquidating) and securities frozen, a view which lingered into the 1920's.[5] Edgeworth indicated the im-

---

1. Diversification can be explained only if the bank is a monopsonist in the market of the highest-yield asset or if it is required by law to carry reserves of low-yield assets.

2. In deference to received literature, the word "risk" is used rather than "uncertainty" since the bank is assumed to know, with certainty, the probability distributions.

3. Cf. pp. 256-259 of K. J. Arrow, T. Harris, and J. Marschak, "Optimal Inventory Policy," Econometrica, Vol. 19, No. 3 (July 1951), pp. 250-272.

4. F. Y. Edgeworth, "The Mathematical Theory of Banking," Journal of the Royal Statistical Society, Vol. 51, Part 1, (March 1888), pp. 113-127.

5. For a review of this revolution in bankers' ordering of relative liquidities, whereby securities became "secondary reserves" and loans frozen assets, cf.

portance of probability to banking through the device of a simple game:

> I have imagined a new game of chance, which is played in this manner: each player receives a disposable fund of 100 counters, part of which he may invest in securities not immediately realizable, bearing say 5 per cent per ten minutes; another portion of the 100 may be held at call, bearing interest at 2 per cent per ten minutes; the remainder is kept in the hands of the player as a reserve against certain liabilities. ... [22 digits are drawn at random every two minutes, and the difference between their sum and their expected sum, 99, is calculated.] The special object of the reserve above mentioned is to provide against demands which exceed that average. If the player can meet the excess of demand with his funds in hand, well; but if not he must call in part, or all, of the sum placed at call, incurring a forfeit of 10 per cent on the amount called in. But if the demand is so great that he cannot even thus meet it, then he incurs an enormous forfeit, 100£ or 1000£.[1]

Unfortunately, Edgeworth then proceeded to solve for the optimum portfolio by a kind of enlightened common sense, claiming that "the calculus cannot indeed, I think, by itself determine what chance of great disaster it might be prudent to incur for the probability of a moderate gain."[2] If however, values can be placed upon the various aspects of this "great disaster," the calculus can do just that.

2. Assumptions

It is uncertainty, in its various guises, far more than anything else which makes the banker's job a difficult one. The important areas of this uncertainty arise because the bank cannot know exactly:

1. How large will be its deposit liabilities at any moment of the future.
2. The market value of the non-matured securities in its portfolio at any moment of the future.
3. What proportion of its borrowers will default the loans which the bank has extended to them.[3]
4. The degree of "frozen-ness" of the loan portfolio at any moment of the future, where this degree depends upon the ability (and, to a certain extent, willingness) of customers to accept refusal of loan renewals.[4]

B. Suviranta, "The Shiftability Theory of Bank Liquidity," Economic Essays in Honor of Gustav Cassel (London: 1933), pp. 623-635.

1. Edgeworth, op. cit., p. 120.

2. Ibid., p. 121.

3. There is also the possibility of default on securities, but Government obligations comprise so large a part of banks' portfolios that this area of uncertainty may be neglected.

4. The nominal maturity distribution of the bank's loan portfolio may have little to do with the actual degree of "frozen-ness" of its loans. A study by the Federal Reserve Bank of Cleveland indicated that continuous borrowing through renewal of short-term loans was quite widespread. While only six per cent of the loans of banks in that district matured in five years or more, 25 per cent of the total dollar amount of loans had been made to borrowers who

While the first element of uncertainty is particularly critical to the bank, the last three areas are clearly not unimportant. If bonds were always marketable at par and loans callable on demand, without possibility of default, the bank could never become illiquid no matter how erratic the behavior of its deposits. The greater the extent to which any or all of these latter three uncertainties exist, the greater becomes the bank's first concern for the future course of deposits. Thus, no one of these four aspects may be properly neglected in a model of bank operations.

The assets which the bank can hold may be divided into three general categories: cash assets, securities, and loans. Since the problem of diversification within each of these portfolios (that is, what types of securities and loans are held) will not be of concern in this paper, each of these categories will be assumed internally homogeneous. "Cash" assets in fact consist of Federal Reserve Bank reserves, vault cash, net balances with other banks, and bills of very near maturity; here no such distinctions will be made, all "cash" being assumed 1) to provide no earnings and 2) to be completely free of risk of capital value change. The category "securities" will be assumed to include a homogeneous group of securities 1) without default risk, 2) readily saleable upon established markets, 3) with maturity date beyond the end of the bank's present planning horizon, and 4) with a fixed coupon per bond per planning period. The distinction between "cash" and "securities," is clearly one of degree and not of kind. The portfolio of an actual bank will invariably consist of a variety of assets in the range from cash to fairly long-term bonds; in this simplified representation, the choice of the bank is narrowed. "Loans" are assumed 1) to be not callable during the planning period, 2) to be not marketable, and 3) to be "shiftable" only to the extent that they are eligible as collateral for borrowing from the Federal Reserve Banks. Thus, the essential difference between "securities" and "loans" is that there is a market for the former so that securities may be readily converted into cash, although at an uncertain price, while loans can be so converted only through the Federal Reserve Bank. Since these are assumed to be the only assets which the bank can hold,[1] it must be true that cash plus securities plus loans equals deposits[2] plus total capital accounts; the bank is assumed to have no liabilities other than deposits, and, of course, no part of the total capital accounts (which will be called simply "net worth") can be withdrawn from the bank.

What has been called the "planning period" is that span of time upon which the bank concentrates all its attention and over which it sets, and does not plan to alter, its asset portfolio. This is obviously unrealistic, for every bank is always planning and re-planning its asset portfolio. Even if the fact of continually maturing securities--which forces the bank to re-plan by automatically replacing securities with cash assets--were removed, as it is in the model, actual banks would make continual changes in their portfolio plans. Nevertheless, it is equally true that portfolios are not planned with the in-

---

had been in debt continuously to the same bank for over five years. Federal Reserve Bank of Cleveland, Monthly Review, September 1956.

1. Non-financial assets comprised less than one per cent of the assets of member banks of the Federal Reserve System in 1956.

2. Deposits are also assumed internally homogeneous, i.e., no distinctions are made between demand and time deposits. The question of the bank's optimal proportion of time to demand deposits is briefly treated in Appendix D.

tention of making frequent changes, and it seems more realistic to assume that the basic portfolio decisions with respect to the fundamental components, cash, securities, and loans, are made fairly seldom and with reference to a sizable span of time. Forcing this flexible procedure into a planning period of fixed length is very simplifying but, it is hoped, not badly distorting.

The choice of this portfolio is assumed to depend entirely upon anticipations concerning the circumstances of the ensuing period and in no way upon past events, past portfolio selections, or expectations of events occurring after the end of the ensuing period. To expose the importance of the assumptions of this last sentence some further amplification is required. First, it is assumed that the bank is not influenced by past events except insofar as these affect its estimates of the future. Thus, if the bank has had an unusual proportion of its loans defaulted in the previous period (the third element of uncertainty), this may induce it to re-value its estimates of such risk but it does not reduce its loans solely on the basis of a "once burnt, twice shy" code of behavior. More relevant to the real world is the connection to the "pin-in" effect. Any such effect is assumed away in the model; previous declines in the security price level (as a result of the second element of uncertainty) cannot cause the bank to carry a greater proportion of securities than it otherwise would desire. Second, past portfolio selections do not influence the current choice (except, of course, through their influence on the bank's appraisal of the ensuing period); this requires that there be leeway in the portfolio at the start of any period. Clearly, cash and securities can be converted into other assets at any time[1] so this assumption really applies only to loans--if all the bank's borrowers "required" renewals of their loans (the fourth element of uncertainty) for the ensuing period, the bank would not be able to reduce its loans, whether it wished to or not. Thus, under this assumption the model can only consider those banks for which the proportion of "required" renewals, while perhaps large, is never so large that the bank cannot start the new period with the exact quantity of loans it wishes to make. Third, the bank knows, or estimates with complete confidence, all the parameters of its environment that are relevant to its portfolio choice for the ensuing period. And fourth, the portfolio of the current period is not affected by expectations of change in the parametric climate of the next period. In short, the assumptions about the "period" are those required to keep the model manageable and static--"once burned, once shy," unitary elasticity of expectations (in the Hicksian sense[2]), and limits to the degree of the fourth area of uncertainty.

The first area of uncertainty, that of the future course of the level of deposits, implies that the bank must be prepared for the possibility that withdrawals exceed additions to deposits over a particular time-span. A net reduction of deposits will always occur over the moment of time during which one depositor makes a withdrawal. Not infrequently, a bank will find its deposit levels declining over a few days or weeks. And it is not impossible that seasonal, cyclical, or secular factors will cause a fall in deposits over longer periods. At the beginning of the planning period, the bank recognizes that its

---

1. The fact that transaction costs of change give the existing portfolio some inertia is assumed to be of little importance at the margin.

2. J. R. Hicks, Value and Capital, 2nd ed., Oxford, Oxford Press, 1946, p. 205. This is not strictly correct since expectations are there assumed single-valued. In this paper, any expectations that are not single-valued are assumed to have a probability distribution which, at every moment of time has an arithmetic mean equal to the current market value.

level of deposits during the period may follow a myriad of possible paths, of which continuous rises or continuous falls are but two. In reality, the complete shapes of the possible paths are implicitly considered in the specification of the bank's asset portfolio. But one aspect of the shape of each of these possible paths is of such great importance to the bank that it is here assumed the only aspect considered by the bank--namely, the lowest point to which deposits fall in each of the paths. For it is at this "deposit-low" of the period that the bank is forced to make the most radical adjustment of its asset portfolio in order to meet the demands of its depositors.[1]

This assumption of sole concern with "deposit-lows" is not in itself sufficient to permit complete neglect of the time-shape of deposit changes since the date of occurrence of any "deposit-low" may still be important to the bank. In the interest of simplicity, this problem of the time-path of deposits will be avoided in the following way. At some point toward the end of the period, deposits will reach their low,[2] at which time the bank makes any asset adjustments required; this perhaps necessitates selling some of its securities and/or borrowing on the collateral of some of its securities or loans.

This is not the place for a full discussion of the complex manner (use of Federal Funds, security sales, Federal Reserve Bank discounts, etc.) in which banks in fact can and do meet the problem of insufficient reserves (i.e., insufficient cash assets). Basically, the process may be simplified into the following stages:

Stage 1. The bank meets net withdrawals from its cash assets as long as it can without drawing these assets down below their minimum required level.

Stage 2. Should the cash assets prove insufficient, the bank sells from its security portfolio, at the going market price, and continues to do so as long as it has securities to sell.[3]

---

1. Of course, if deposits rise throughout the period, the "deposit-low" is zero and the bank need make no adjustments as far as meeting withdrawals is concerned. Thus, the "most radical" adjustment may well be no adjustment at all. This "most radical" adjustment may also be slight if the bank has access to short-term borrowings (such as Federal Funds or the discount window) and is not reticent to use them for short periods in hope of improvement in its deposit position. In this case, end-of-period deposits may be a more important variable than the "deposit-low" of the period. The conservative bias that the "deposit-low" assumption gives the portfolio is indicated in Appendix F.

2. If deposits, on the average, should rise during the period, the "deposit-low" will probably not be much, if at all, below zero and will probably occur toward the beginning of the period. Since little asset readjustment is required in this case and since we neglect the net rise in deposits that follows, the assumption that the "deposit-low" occurs toward the end of the period is innocuous.

3. For simplicity, it is assumed that the banks sell, rather than borrow from the Federal Reserve Banks on the collateral of securities. Given bankers' dislike of debt and the fact that interest charges would probably exceed transaction costs of selling and later repurchasing, sales rather than borrowing would occur in the world postulated by the model, that is, a world of no "pin-in" effects and unitary elasticity of expectations of bond prices.

Stage 3. Should the sale of all its securities also be inadequate to meet the deposit depletions, the bank borrows from the Federal Reserve Bank on the collateral of its outstanding loans.[1] This it continues to do as long as necessary or until its stock of such collateral is exhausted.

To these three stages, a fourth might be added: should all its assets be converted, to the greatest extent possible, into means of payment and still be insufficient to cover deposit withdrawals, Edgeworth's "great disaster" would occur par excellence--the bank would then be in the throes of a liquidity crises beyond its ability to handle. At the very least, it would have to call for exceptional aid from the Federal Reserve System; it might be forced to close its doors, and it might find itself insolvent as well. However, it will be assumed that such a "Stage 4" is so costly to the bank that no bank's optimum portfolio permits any possibility of this occurrence. Such a result may (and, in fact, does, if seldom) occur, but it could only happen, by the assumptions here, through a misestimation by the bank of the parameters of its operations.[2]

The method of treatment of the four areas of uncertainty may now be more accurately specified.

1. The size of deposits at any moment of the future. Although the bank expects (in the probability sense) deposits to stay at their start-of-period level, it recognizes that they may fall or rise. The relevant distribution function is that relating each of the various "deposit-lows" during the ensuing period to the probability of its occurrence. If the random variable, u, is defined to be the "deposit-low" as a fraction of initial deposits, the distribution of u may be defined only over the range, zero to unity. For simplicity, the frequency distribution of u, $f(u)$, is assumed to be a linearly increasing function of the amount by which u exceeds s, where s is the smallest "deposit-low" (as a fraction of initial deposits) to which the bank assigns a non-zero probability (and clearly, $0 < s < 1$).[3] Since the cumulative of $f(u)$ must equal one, specification of s is sufficient to determine:

$$f(u) = \frac{2(u - s)}{(1 - s)^2}$$

2. The market value of its securities at any moment in the future. A "unit" of securities is defined as a dollar's worth at the market prices prevailing at the start of the planning period; this "unit" carries a coupon paying g dollars per "unit" per period where it is assumed, without undue restriction, that $0 < g < 1$. The market price at the end of the period may be written as $(1 + w)$ where w, the change in security prices during the period (absolute

---

1. Alternatively, one may think of this process as straightforward rediscounting in the traditional, if in fact little used, manner.

2. It is possible that some bank may be forced to accept the possibility of such a "Stage 4" if its net worth were low and its lowest possible "deposit-low" very near zero. Of course, the bank could meet this situation by holding very large cash assets, but this may be so unprofitable as to induce it to accept the possibility of "Stage 4." Such a bank is not considered in this paper.

3. This "triangular" distribution is assumed because it is believed to be the best simple approximation to the actual distribution of banks' "deposit-lows." For a theoretical derivation of the distribution, see Appendix A. Alternatively, a uniform distribution of "deposit-lows" is considered in Appendex E.

and percentage), is assumed to be uniformly distributed over the range, -a to a ($0 < a < 1$).[1] Since the "deposit-low" occurs toward the end of the period, the price of securities sold at the moment of the "deposit-low" may also be considered to be $(1 + w)$.

3. The proportion of loans defaulted. This aspect of bank uncertainty will be most summarily treated, not because it is felt to be unimportant to a complete theory of bank operations but because its basic effects upon the bank's portfolio can be seen in the present model without complex treatment. It is assumed that the bank charges a pure interest rate of e per dollar of its loans; it then adds to this rate some amount according to the default risk which just suffices to insure that the bank will not lose through defaults in the long run. The final "gross" rate is e' (where $0 < e < e' < 1$), but we shall here concern ourselves only with the bank's earnings net of default.[2]

4. The degree of intra-period "frozen-ness" of the loan portfolio. The fear on the part of the bank that it may not be able in an emergency to reduce its loans sufficiently, even over several periods, means that any debt incurred to help meet deposit depletions may well be long-term debt, a position which bankers dislike.[3] It is because of this that Stage 2, sales of securities, is assumed to precede Stage 3, borrowing on loan collateral, in the process of meeting deposit depletions. For the same reason, the cost of such borrowing in Stage 3 may be interpreted to include not only the charge

---

1. Unfortunately, fixing the distribution of w so that its expected value is zero implies that the expected value of the equivalent distribution of interest rates is greater than zero, in contradiction of the assumption of static expectations with respect to interest rates. In the case of consols the expected value of the distribution of changes in the interest rate is:

$$g\left[\frac{a^2}{3} + \frac{a^4}{5} + \frac{a^6}{7} + \dots\right]$$

which is sufficiently near zero for small values of g and a that this contradiction is not serious. This results from the property of the number system, whereby the average of the reciprocals does not equal the reciprocal of the average.

2. It must always be remembered, however, that to the extent that the bank is worried about the time-path of defaults or the default rate is positively related to the quantity of loans, the present model will overstate the amount of loans which the bank will desire to make.

3. Much nebulous writing has appeared on this subject, but the bankers' aversion seems real enough, probably basically deriving from their fear that heavy indebtedness will have adverse effects upon their relations with depositors, borrowers and correspondent banks. The view is not without its dissenters, however; for example, see A. Murad, "The Ineffectiveness of Monetary Policy," Southern Economic Journal, Vol. 22, No. 3 (January 1956), pp. 339-351. Of bankers' supposed aversion to steady borrowing, Murad says (p. 346):

> It may be that bankers feel that way or say that they feel that way, but they certainly do not act that way. Whenever they have reserve deficiencies they borrow and if necessary remain for years in debt to the Federal Reserve banks.

As a general phenomenon, this last sentence is open to great doubt, for the fact of increasing or large aggregate indebtedness is not proof of a decreasing or small antipathy toward permanent indebtedness.

of the Federal Reserve Bank but also a subjective "cost" of being in what may prove long-term debt.

While use of the "discount window" is a privilege and not a right, no Federal Reserve Bank would refuse to extend advances to a bank which found itself unable to cover exceptionally large deposit withdrawals without such aid. The only questions before the bank are, then, how much borrowing could they do on the basis of their total loan portfolio and how much would it cost. If the bank gets into Stage 3, it can take a typical dollar's worth of its loans to the Federal Reserve Bank and receive an advance of $(1 - m)$ dollars, where m, which might be labelled the "excess-collateral rate," is, of course, between zero and unity. On this advance, the borrowing bank is charged an interest cost which, it is here assumed, is different from the real "cost" because of bankers' dislike of such debt. There are many ways in which such a "cost" might be handled, but the one to be assumed in this paper is that the real "cost," q, of such borrowing is to some extent greater than the interest charge, where $0 < q < 1$.[1]

Cash assets are of two types, those required to be held as reserves in the Federal Reserve Bank and those which the bank holds (in various forms) in excess of these requirements. The amount of the former at any moment of time must be a specified fraction of the bank's deposit liabilities. While, in fact, a rise in reserve requirements usually results in the lowering of the amount of other cash assets which the bank feels it requires, it is here assumed that the amount of cash assets other than required reserves is also a specified fraction of its current deposit liabilities. Thus, the "required" amount of cash assets can be written as a fraction, k, of the bank's deposit liabilities, where k is somewhat larger than the reserve requirement ratio. Most banks would meet these requirements almost continually, but as the model is set up, it need only be prepared to meet them at the moment of the "deposit-low" to be sure of having a sufficient amount at every other moment of the period. Assuming a fixed fraction of cash assets in this fashion means that Stage 1 is not possible. But this assumption is not as restrictive as it might seem at first since the excess of cash assets over k would probably be very small unless g were near zero and/or a extremely large.[2]

Each of the balance sheet items will be written as a fraction of start-of-period deposits--the fraction of cash assets being k, of securities, B, of loans, L, and of net worth, N. B and L are variables under the control of the bank, while N is assumed previously determined and unalterable at least over the ensuing period.

It remains only to fix the criterion by which the bank balances its portfolio between possible gains and losses. One often stated by bankers themselves is that they minimize the probability of losses (or, in reverse, maximize the probability of some gain); but this implies that the portfolio be prepared to meet any possible deposit reduction out of cash assets, while in fact

---

1. If there were no addition of a subjective "cost" and the bank knew that it could repay its debt in exactly one period, q would be equal to the Federal Reserve Bank discount rate (per period). To the extent that there is a subjective element or such last-resort borrowing is felt to be of longer duration, q may be well above the discount rate.

2. The real-world analog to the parameter g is the difference between long-term security rates and the bill rate. Not infrequently, this difference is very slight, but such times have little relevance here for the assumption of static expectations concerning future interest rates is then almost certainly violated.

banks do incur an unnecessary, if small and profitable, risk of losses. A variant of the above is the minimization of the probability of incurring losses within the constraint of a reasonable expected profit. Such a criterion is rejected on the grounds that setting the definition of "reasonable" is more important than the minimization process that follows. A criterion advanced by recent portfolio theory[1] is that some point is chosen, according to the selector's preferences, on the frontier (or locus) of the maximum expected return for every possible variance of return. This criterion was introduced because it was useful in explaining diversification; in the present model, the variable, variance of return, is not needed to explain diversification and so, for simplicity, will not be included in the text.[2] Here, the bank is assumed to choose its asset portfolio so as to maximize its expected additions to net worth during the period. Thus, it maximizes its expected additions to net worth function with respect to one of the asset variables, B and L, the other being then determined by the accounting identity:

(1) $1 + N = k + B + L$

In summary, the symbols to be used in the paper are:

Variables:

1. B, securities as a fraction of initial deposits.
2. L, loans as a fraction of initial deposits.

Random elements:

3. u, the "deposit-low" of the period, as a fraction of initial deposits. u is defined over the range, $0 < s < u < 1$, by the distribution, $f(u) = \frac{2(u - s)}{(1 - s)^2}$.
4. w, the change between the start of the period and the "deposit-low" (and the end) of the period in the market price of securities. w is defined over the range, $-a < w < a (0 < a < 1)$, by a uniform distribution, $f(w) = \frac{1}{2a}$.

Parameters:

5. N, net worth, unchanging, as a fraction of initial deposits.
6. g, the coupon per dollar's worth of securities (at initial market prices); $0 < g < 1$.
7. e, the earning rate on loans (net of default risk); $0 < e < 1$.
8. k, the amount of cash assets which the bank holds, as a fraction of current deposit liabilities; $0 < k < 1$.
9. q, the "cost," both actual and subjective, of borrowing a dollar during the period from the Federal Reserve Bank; $0 < q < 1$.
10. m, the "excess-collateral rate." A dollar of loans as collateral enables the bank to borrow (1 - m) dollars from the Federal Reserve Bank (the latter acting in its capacity of "lender of last resort"); $0 < m < 1$.

---

1. Cf. H. Markowitz, Portfolio Selection (New York: 1959), Part IV; J. Tobin, "Liquidity Preference as Behavior Towards Risk," Review of Economic Studies, Vol. 25, No. 2 (February 1958), pp. 65-86; and I. O. Scott, "The Availability Doctrine: Theoretical Underpinnings," Review of Economic Studies, Vol. 25, No. 1 (October 1957), pp. 41-48.

2. The implications of a bank preference map which is a function of both expected gain and variance of gain are drawn, for a simplified version of the model, in Appendix B. Also, some inference about the change in results which this would cause is made there.

3. Structure of the Model

The amount of profit which the bank makes during the period will clearly depend upon the "stage" into which the "deposit-low" forces it, and, for Stage 3, upon how far into that stage it goes. The expected addition to net worth (ΔN) for each stage is:

Stage 2. (Securities sales are required to handle the "deposit-low.")

$$(2) \qquad \Delta N = gB + wB + eL$$

The profit is composed of: gB, earnings on securities; wB, capital gains or losses on securities; and eL, earnings (net) on loans.

Stage 3. (Borrowing on loan collateral is required.)

$$(3) \qquad \Delta N = gB + wB + eL - q(1 - m)x$$

where x is the amount of loans put up as collateral (and $(1 - m)x$ the amount borrowed) and

$$(4) \qquad (1 - m)x = (1 - k)(1 - u) - (1 + w)B$$

The profit is composed of: gB, securities earnings; wB, capital gains or losses on the (completely) sold securities; eL, earnings on loans; and $q(1 - m)x$, the cost of the bank's borrowings.

According to the values assumed by the random variables, u and w, the bank finds itself in one of these two stages. The ranges of u and w which bring about each stage are:

Stage 2. This may occur in either of two ways:

i) $s < u < 1$; $-1 + \frac{(1-s)(1-k)}{B} < w < a$. It may be possible that the most extreme (conceivable) deposit depletion can be met through bond sales alone, provided that the price of securities rises sufficiently (or falls sufficiently little) during the period.

ii) $1 - \frac{(1+w)B}{1-k} < u < 1$; $-a < w < -1 + \frac{(1-s)(1-k)}{B}$. If securities prices do not rise enough, only a certain degree of possible deposit withdrawals can be handled by means of bond sales alone.

Stage 3. $s < u < 1 - \frac{(1+w)B}{1-k}$; $-a < w < -1 + \frac{(1-s)(1-k)}{B}$. Any deposit depletion too extreme to be met by securities sales alone can be met by bond sales plus borrowings on loan collateral.

These three cases are exhaustive since we have already excluded the possibility of a Stage 1, the running down of cash assets to meet withdrawals, and of a Stage 4, where even borrowings on all the bank's loan collateral are insufficient to cope with the deposit losses. But the fact that the three possibilities listed above cover the entirety of the ranges of w and u does not automatically imply that every stage is relevant for every bank--for example, the net worth of a bank may be so high that it is able to cover any conceivable deposit depletion by bond sales alone, even if the price of bonds drops to their lowest conceivable level. In technical terms, for a stage to be possible of occurrence, the lower limits of both u and w (for that stage) must indeed be lower than the upper limits. If all three of the cases cited above--Stage 2(i), Stage 2(ii), and Stage 3--are considered relevant to the bank, then the following assumption about the size of the bank's securities holdings is being made implicitly:

$$(5) \qquad \frac{(1-s)(1-k)}{1+a} < B < \frac{(1-s)(1-k)}{1-a}$$

for if the left-hand inequality does not hold, Stage 2(i) does not exist; and if the right-hand inequality does not hold, Stage 2(ii) and Stage 3 do not exist.

The bank for which Stage 3 is not even a remote possibility is not only rare but uninteresting, so there should be few qualms about assuming the right-hand inequality. But the left-hand inequality is not so easily handled; a taxonomic approach would construct the model for both directions of the inequality sign, but here only that one which is felt most closely to describe reality will be extensively treated.[1] In a world where banks do not expect extremely large bond price fluctuations and do make a significant amount of loans relative to their net worth and lowest possible "deposit-lows," most banks are probably not able to cover their lowest conceivable "deposit-lows" by means of securities sales alone, even if bond prices rise to as high a level as is considered possible. Thus, in what follows, it is assumed that Stage 2(i) is not a possibility, or, in other words, that inequalities (5) may be replaced by:

$$(6) \qquad 0 < B < \frac{(1-s)(1-k)}{1+a}$$

4. Determination of the Optimum Portfolio

The stage in which the bank finds itself is determined, as follows, by the value assumed by the random variable, u :

Stage 2(ii). $1 - \frac{(1+w)B}{1-k} < u < 1 \; ; \quad -a < w < a \, .$

Stage 3. $s < u < 1 - \frac{(1+w)B}{1-k} \; ; \quad -a < w < a \, .$

The expression for the expected addition to net worth, $E[\Delta N]$, is:

$$(7) \qquad E[\Delta N] = \int_{w=-a}^{a} \int_{u=s}^{1} [gB + wB + eL] \left[\frac{u-s}{a(1-s)^2} \right] du \, dw$$

$$+ \int_{w=-a}^{a} \int_{u=s}^{1-\frac{(1+w)B}{1-k}} [-q(1-k) + q(1+w)B + q(1-k)u] \left[\frac{u-s}{a(1-s)^2} \right] du \, dw$$

which becomes, after integration,

$$(8) \qquad E[\Delta N] = gB + eL - \frac{q}{3(1-s)^2(1-k)^2} [(1-s)(1-k)-B]^3$$

$$- \frac{a^2 q B^2}{3(1-s)^2(1-k)^2} [(1-s)(1-k)-B]$$

L may be eliminated as a variable by means of the accounting identity (1); then the derivative of $E[\Delta N]$ with respect to B is:

1. Enough has been worked through for the other case to indicate that the results are similar. See Appendix C for a brief treatment of the bank which is able, for some conceivable level of bond prices, to meet the worst conceivable "deposit-low" without recourse to borrowing.

$$(9)\qquad dE[\Delta N]/dB = g - e + \frac{q}{(1-s)^2(1-k)^2}\,[(1-s)(1-k) - B]^2 - \frac{a^2 q B}{3(1-s)^2(1-k)^2}\,[2(1-s)(1-k) - 3B]$$

and the second derivative:

$$(10)\qquad d^2E[\Delta N]/dB^2 = \frac{2q}{(1-s)^2(1-k)^2}\,[-(1-s)(1-k)(1+\tfrac{a^2}{3}) + (1+a^2)B]$$

If B is at its lowest permissible level (i.e., zero)[1] the first derivative, $dE[\Delta N]/dB$, is positive, and, if B is at its highest value (i.e., $\frac{(1-s)(1-k)}{1+a}$), the derivative is negative, provided that the following inequalities hold:[2]

$$(11)\qquad \frac{2a^2(2-a)}{3(1+a)^2} < \frac{e-g}{q} < 1$$

If inequalities (11) hold, there exists a regular maximum and the optimum fraction of assets in securities can be found by setting the derivative (9) equal to zero. Solving this quadratic equation in B yields:

$$(12)\qquad B = \frac{(1-s)(1-k)(1+\frac{a^2}{3})}{1+a^2}\left\{1 - \sqrt{1 - \frac{(1+a^2)(1-\frac{e-g}{q})}{(1+a^2/3)^2}}\right\}$$

Much of the complexity (or rather simplicity!) of the form of equation (12) results from the particular functional representation of the distribution of the random variables, u and w; but there are three interesting properties of (12) which are not dependent upon the choice of distribution functions:

1. The fact that (1-s)(1-k) enters in linear fashion. In words, this quantity is the fraction which does not need to be held in cash assets (1-k) of the largest conceivable loss in deposits (1-s). The rest of equation (12), involving the parameters, a, e, g, and q, serves to fix the quantity of securities as a fraction of this term, (1-s)(1-k). Thus, the determination of the optimum portfolio can be divided into two problems, first, the calculation of the maximum amount of securities which the bank would ever need to sell to meet deposit losses (on the assumption that bond prices do not change), and second, the decision as to what fraction of this amount the bank will actually hold (which will depend upon the various earning and borrowing cost rates as well as the expected fluctuation in security values).

2. The manner in which the earning and borrowing cost rates enter the equation. By means of the single expression, (e-g)/q, the bank measures the relative advantage of loans vis-a-vis securities, the advantage being greater the larger is the difference in earning rates and/or the lower is the

---

1. See inequalities (6).

2. See Appendix C. If the left-hand inequality does not hold, the present model is inapplicable and that treated in Appendix C becomes the relevant one. Unless very high values of a are considered, however, the expression on the left will be very close to zero. If the right-hand inequality does not hold, there is a corner maximum, with no securities entering into the portfolio.

cost of borrowing on loan collateral. It is interesting to note that, if the bank is to include any securities at all in its optimal portfolio, it is not necessary that the borrowing cost be greater than the loan earning rate (i.e., a "penalty" rate is not essential), but only that it be larger than the difference between the earning rate on loans and that on securities.

3. The fact that N, the net worth of the bank, plays no part in the determination of the optimum quantity of securities. This independence between optimum B and N implies that any change in N induces a change in the bank's loans of the same amount and direction. This conclusion is not surprising if one recognizes that net worth, from the viewpoint of the bank's liquidity problems, can be treated as a deposit liability with no possibility of withdrawal.

What is really interesting about equation (12) is not the level of B, and hence of L, but rather the way in which the optimum holdings of these assets vary as a result of changes in the different parameters. The first and second partial derivatives of B with respect to the various parameters are given in the table below.

TABLE 1

		$\partial B/\partial x$	$\partial^2 B/\partial x \partial y$ where					
			$y = a$	$y = q$	$y = g$	$y = e$	$y = k$	$y = s$
where	$x = s$	-	$\pm^1$	-	-	+	+	0
	$x = k$	-	$\pm^1$	-	-	+	0	
	$x = e$	-	$\pm^2$	+	-	+		
	$x = g$	+	$\mp^2$	-	+			
	$x = q$	+	$\mp^2$	-				
	$x = a$	$\mp^1$	$\mp^3$					

1. upper or lower sign holds according as: $\frac{e - g}{q} \gtrless \frac{1}{9}$

2. upper or lower sign holds according as: $\frac{e - g}{q} \gtrless \frac{3 - 2a^2}{9}$

3. evaluated at $a = 0$; upper or lower sign holds according as:

$$\frac{e - g}{q} \gtrless \frac{1}{9}$$

(see Appendix E for the signs when a uniform distribution of "deposit-lows" is used.)

A similar table of partial derivatives could be constructed for the changes in optimum holdings of loans, but this is not necessary; since, by accounting identity (1), $L = 1 + N - k - B$, the first and second partial of

L with respect to any parameter (except k) is simply the negative of the relevant partial of B. It can be shown that an increase in k decreases L as well as B, and that the second derivative $(\partial^2 L/\partial k^2)$ is zero.

One could find quantitative estimates of these derivatives by assuming particular parameter values, but more generally we can plot the value assumed by $\frac{B}{(1-s)(1-k)}$ (written hereafter as B') for all possible values of a and of the composite parameter, $(e - g)/q$. This is done, for various fixed levels of $(e - g)/q$, in Figure 1. The dotted line is the border above and to the right of which Stage 2(i) becomes a possibility.[1]

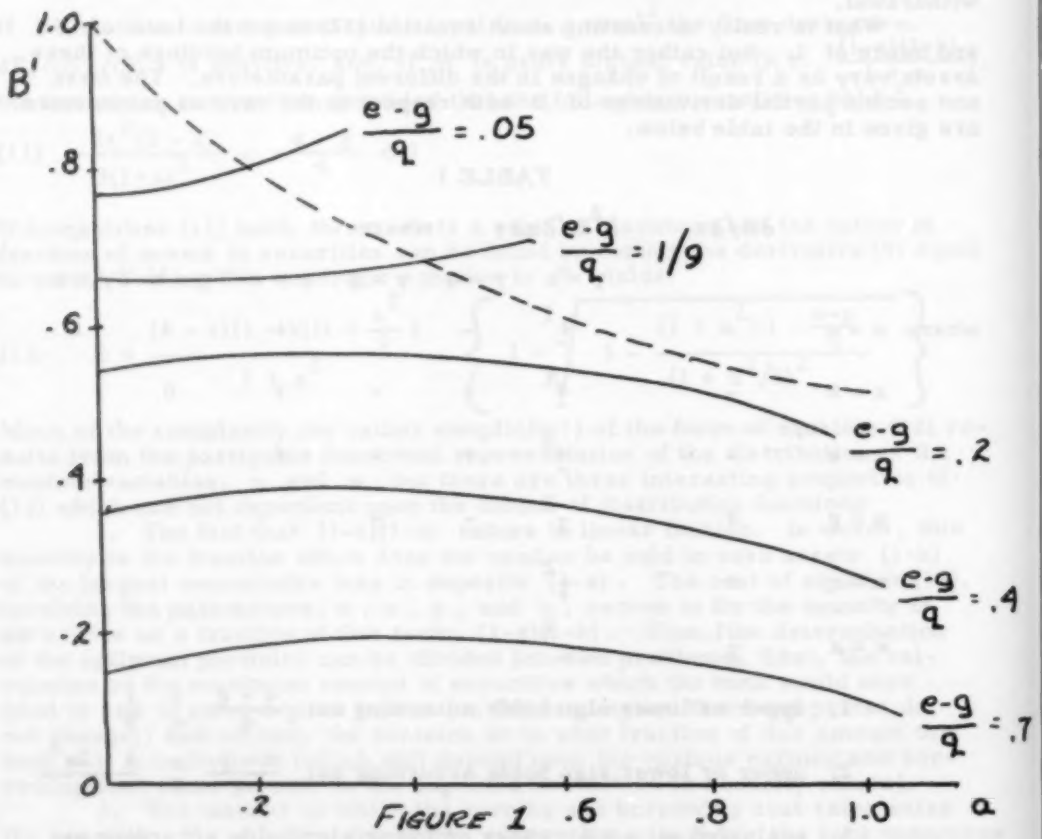


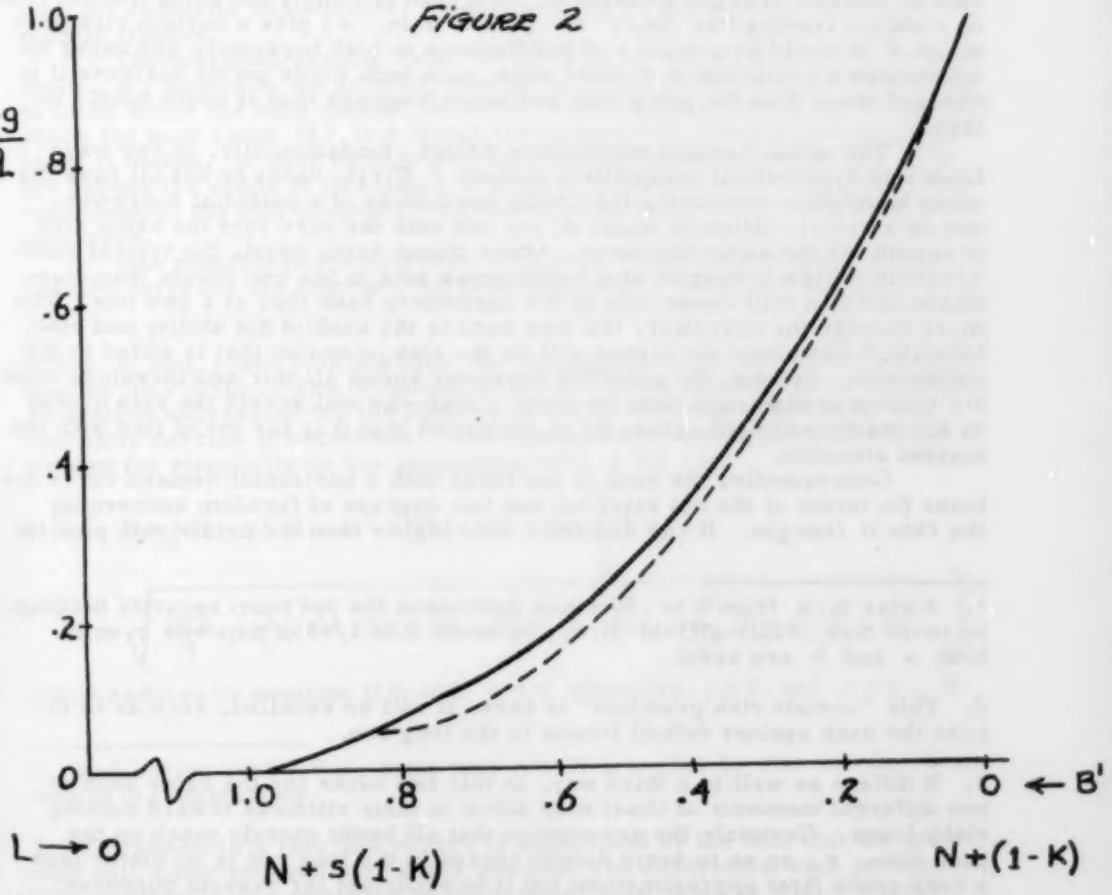
FIGURE 1

1. Although that region is neglected in Figure 1, numerical examples based on Appendix C indicate that the curves could be smoothly extrapolated into the State 2(i) area without much, if any, error.

The most obvious lesson of Figure 1 is that the effect of changes in the anticipated fluctuation in security prices is uncertain, with respect to both direction and magnitude. When $(e - g)/q$ is in the neighborhood of $1/9$ or of unity (i.e., when B' is in the neighborhood of .67 or zero), changes in the parameter, a, have almost no effect upon the composition of the optimum portfolio. The farther $(e - g)/q$ is from these critical values, the greater will be the effect on the portfolio of a. If $(e - g)/q$ is less than $1/9$, greater certainty about the course of future bond prices will induce the bank to hold less securities; while if $(e - g)/q$ is between $1/9$ and unity, greater bond-price certainty will induce larger holdings of securities.

Lest Figure 1 give the impression that the value of a is a critical determinant of the portfolio composition, another diagram, Figure 2, is included which relates B' to $(e - g)/q$ for two very different values of a. The solid line shows the relation at $a = 0$ and the dotted line at $a = \frac{1}{2}$.[1]

FIGURE 2



1. The dotted line is not continued to the point where $(e - g)/q$ is zero because for the very low values, Stage 2(i) becomes possible. For some values of a, as Figure 1 shows, the dotted line will cross the solid one for low values of $(e - g)/q$.

It will be seen from Figure 2 that a has no more than a very marginal effect upon the portfolio,[1] and that the important parameters for the division of the portfolio between loans and securities are, not surprisingly, the difference in earning rate between loans and securities, (e - g) , and the cost of borrowing, q . It is the influence of these latter parameters (as well, of course, as s and k) with which the rest of this paper is concerned; considerable simplification will henceforth be achieved by the assumption that a is zero. The first step will be to drop the unrealistic assumption that e is a constant, unaffected by the quantity of loans which the bank makes.

5. Imperfect Competition in the Loan Market

In order to conceive of the bank as, to some degree, a monopolist in its loan market, the meaning of the loan demand curve must be analyzed. As long as banks are a homogeneous group, each of which have available the same information concerning the credit-worthiness of every potential borrower, the rate of interest charged a customer for a loan is simply the going market rate on riskless lending (the "pure" or "prime" rate, e) plus a certain risk premium.[2] It would be a matter of indifference to both borrowers and banks to which bank a particular borrower went; each bank would get no business if it charged more than the going rate and more business than it could handle if less.

The actual banking mechanism differs, fundamentally, in two ways from this hypothetical competitive system.[3] First, banks do not all have the same knowledge concerning the credit worthiness of a potential borrower, and as a result, different banks do not add onto the pure rate the same risk premium for the same borrower. Other things being equal, the typical businessman is able to borrow at a lower gross rate in his own locale than elsewhere and at a still lower rate at his customary bank than at a new one. The more strange the borrower, the less sure is the bank of his ability and reliability,[4] and hence the higher will be the risk premium that is added to the prime rate. Second, the potential borrower knows all this and therefore tends not to shop around each time he seeks a loan--he will accept the rate quoted by his traditional bank unless he is convinced that it is far out of line with the market situation.

Consequently, the bank is not faced with a horizontal demand curve for loans (in terms of the net rate) but has two degrees of freedom concerning the rate it charges. It can demand a rate higher than the prime rate plus its

---

1. A rise in a from 0 to .50 never decreases the optimum security holdings by more than .022(1-s)(1-k) (i.e., by more than 1/40 of deposits even if both s and k are zero).

2. This "certain risk premium" is here, it will be recalled, such as to insure the bank against default losses in the long run.

3. It differs as well in a third way, in that two banks (or the same bank at two different moments of time) may differ in their attitudes toward making risky loans. Certainly the assumption that all banks merely mark up the pure rate, e , so as to avoid default losses in the long run is no better than a very crude first approximation; but it is sufficient for present purposes.

4. And the more expensive it is to ascertain. If the bank has been dealing with the borrower for a long time, it does not need to incur the costs of careful credit investigation.

proper estimate of the risk premium and not lose all its customers because even this gross rate will be lower than many of its borrowers could get elsewhere. Moreover, even those of its borrowers who could do better by taking a higher risk premium but a lower gross rate at another bank are not likely to realize this immediately.[1]

Just as the bank does not lose all its loan business by raising its pure rate above the going market rate, so also does it not gain an infinite amount of new loan demand by undercutting the going rate. For it would need to undervalue its risk premiums and overcome other bank-borrower inertias in order to gain the new business.

All the above assumes that the same pure rate must be charged to all borrowers. To the extent to which the bank can discriminate between borrowers,[2] it improves its situation (at least until its customers find out). At the extreme of perfect discrimination, the demand curve for the bank's loans in terms of the pure rate represents not its average but its marginal earning rate (net of default) per dollar of loans. While the following argument applies, with the requisite adjustments, equally well for this case, it will not be explicitly treated.

The marginal earning rate, or marginal revenue, of loans (net of default), e, should in the general case be written as a function of the amount of loans which the bank makes; however, since the fraction of initial deposits which the bank lends (L) is a linear transformation of the dollar amount of loans it makes, e may equally well be considered a function of L. It is far beyond the scope of this paper to predicate the details of this functional relationship; the only property it is safe to assume is that e declines as L increases.

One possible procedure from this point would be to hypothesize a specific form for the function; for example, that e is linear in L:

$$e = h - jL \tag{13}$$

where h and j are both positive; the average earning rate on loans would then be $(h - \frac{1}{2}jL)$. The technique of Section 4 can again be used to derive an explicit expression for the fraction of initial deposits which the bank optimally carries in securities. The monopolistic analog of equation (12), written for simplicity on the assumption that $a = 0$, is:

$$B = \frac{j}{q}(1-s)^2(1-k)^2 + (1-s)(1-k)\left\{1 - \sqrt{\frac{h-g}{q} + \frac{j}{q}[(1-s)(1-k) - 2(1+N-k)] + \frac{j^2}{q^2}(1-s)^2(1-k)^2}\right\} \tag{14}$$

which reduces to equation (12) with $a = 0$ whenever $j = 0$ and $e = h$. It

1. And having realized it, many may not wish and/or be able to take advantage of it immediately.

2. The probability that this occurs is augmented by the fact that the gross rate will differ between borrowers anyway, and differential risk premiums help to disguise the existence of differential pure rates as well. Furthermore, the fraction (if any) of the loan which the bank insists (or strongly suggests) be retained in the borrower's deposit may vary among borrowers--and this practice is essentially nothing more than a rate increase.

is interesting to note three important differences between equations (12) and (14): first, the term $(1 - s)(1 - k)$ no longer enters as a mere proportioning factor in the determination of optimum B; second, j, slope of the marginal revenue from loans function, is an essential element in the determination of B; and third, N now has an effect upon the size of B as well as L.

One could now proceed, as in the previous section, to derive the various properties of the derivatives of equation (14), but this will not be done partly because of the mathematical complication, partly because the results would have validity only for the special case of a linear demand for loans function,[1] but mostly because it is unnecessary. Traditional economic theory suggests that, if we know the marginal revenue from loans function, we can deduce the relevant implications of the market if only we can discover the marginal cost function. And such a function we have already found--implicitly--in the derivation of equation (12).

6. The Marginal Cost of Loans

Equation (12) is the expected profit maximizing relation between the optimum quantity of bonds and the parameters of the model. By means of the accounting identity (1), the following equation for the optimum amount of loans may also be found:[2]

$$(15) \qquad L = N + s(1 - k) + (1 - s)(1 - k)\sqrt{\frac{e - g}{q}}$$

So long as all the parameters on the right-side of equation (15) are considered unalterable constants, L is simply a function of these parameters, and there is no need to go further. But in Section 4, the possibility was introduced that the marginal earning rate of loans depends upon the quantity of loans made. When this is so, it is better to view equation (15) as a relationship between two variables, the optimum proportion of loans, L, and the marginal loan earning rate, e.[3] Then equation (15) may be put into a form more readily seen as a marginal cost function merely by algebraic manipulation of e to the left-side of the equation:

$$(16) \qquad e = q\left[\frac{L - N - s(1 - k)}{(1 - s)(1 - k)}\right]^2 + g$$

In effect, equation (16) is the bank's "marginal cost of loans" schedule in that it shows what the marginal revenue (or marginal earning rate) of loans must be if the bank is to make any given amount of loans.

---

1. If e were made a function of powers of L higher than the first, the maximization equation would involve third (or higher) powers of B, and hence B could be written as an explicit function of the parameters. Implicit differentiation would, of course, still be possible.

2. Throughout this section, too, the parameter, a, is assumed zero for simplicity.

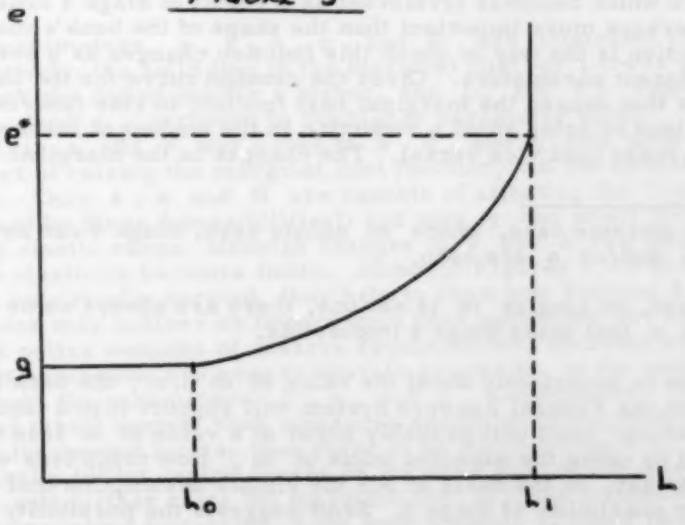
3. The e which enters in equations (12) and (15) is the marginal and not the average earning rate on loans. This can be seen by looking at equations (8) and (9): The e appears as a result of the differentiation of the term (eL), and this derivative of the total revenue of loans is the marginal revenue of loans.

The range of L over which equation (16) is relevant is, however, limited. First, L must be less than (or equal to) the total of the non-cash asset portfolio, $(1 + N - k)$, since B has then taken its smallest possible value, zero. Once L has attained this limit, the marginal cost of loans, e, will equal $(g + q)$, and no further increases in e can induce the bank to augment its loan portfolio. Hence, "capacity" limitations imply that the marginal cost of loans curve becomes vertical (i.e., perfectly inelastic with respect to e) at a value of L equal to $(1 + N - k)$. The second limitation is less obvious. Equation (12) and hence equation (16), has been derived on the assumption that there is a possibility of Stage 3 (where borrowing on the collateral of loans is required if the bank is to meet the withdrawals of its worst possible "deposit-low"). But this assumption is violated if L becomes less than $[N + s(1 - k)]$. Thus equation (16) is the relevant marginal cost function only over the range,
$[N + s(1 - k)] < L < [1 + N - k]$.

No elaborate theory is needed to understand the bank's actions in the range, $0 < L < [N + s(1 - k)]$, for then the bank faces no potential liquidity problems--sales of securities will always suffice to cover any conceivable amount of withdrawals. The bank's only concern is its earnings, and it will maximize these by making loans as long as the marginal earnings rate on loans exceeds the earning rate on securities; once e equals g, the bank will expand loans no further, holding the remainder of its non-cash asset portfolio in securities. The marginal cost of loans function is, therefore, horizontal (i.e., infinitely elastic with respect to e) at the level of g over the range of L less than $[N + s(1 - k)]$.

This marginal cost of loans curve is shown in Figure 3, where $L_0 = [N + s(1 - k)]$, $L_1 = [1 + N - k]$, and $e^* = g + q$. The function has

FIGURE 3



three principal regions: one of infinite elasticity for low values of L; one of zero elasticity once all possible loans have been made; and one of intermediate values of L where the elasticity declines continuously from infinity to zero as L increases. It is true that the resemblance of this to the traditional manufacturing cost curve is slight--both are equated to marginal revenue to deter-

mine the optimum "output"--but this follows from the fact that manufacturing cost curves are only indirectly aligned to opportunity costs. As long as L is less than $[N + s(1 - k)]$, the banks loan costs are simply the opportunity costs of an alternative "output," i.e., holding securities. If L is large enough to make Stage 3 possible, to these opportunity costs is added an illiquidity-incurring cost of increasingly greater size as L rises.

Throughout this paper, we have neglected the possibility of a Stage 4 (where sale of all securities and borrowing on the collateral of all loans are insufficient to cover the withdrawals of the worst conceivable "deposit-low"). This neglect was justified on the grounds that no bank would ever choose a portfolio that permitted any possibility of so fearful an occurrence. But it is possible that Figure 3, as drawn, violates this assumption. Stage 4 emerges as a possibility if L is greater than $[N + s(1 - k)]/m$, where m is the "excess-collateral rate" required by the Federal Reserve Bank for the bank's borrowing on the collateral of its loans. For very small values of m, it is clear that this consideration will be irrelevant,[1] but there are also values of m large enough to induce the Stage 4-avoiding bank to cease its loan expansion, regardless of the marginal earning rate, before loans comprise the entire non-cash asset portfolio. This Stage 4 constraint becomes potentially operative if

$$(17) \qquad m > \frac{N + s(1 - k)}{1 + N - k}$$

The bankers talk of being "loaned-up" while their portfolios still carry some securities is perhaps partial evidence that some such Stage 4 restriction generally does occur.[2] If so, the marginal cost curve becomes vertical not at L_1 but at L_2 (where loans equal $[N + s(1 - k)]/m$);[3] this is shown in Figure 4, where the dotted curve indicates the part of the Figure 3 marginal cost curve which becomes irrelevant as a result of Stage 4 considerations.

Perhaps more important than the shape of the bank's marginal cost of loans function is the way in which this function changes as a result of changes in the different parameters. Given the demand curve for the bank's loans, any factor that causes the marginal cost function to rise (and/or shift to the left) will tend to bring about a reduction in the amount of loans which the bank wishes to make (and vice versa). The changes in the marginal cost function,

---

1. In the extreme case, where m equals zero, Stage 4 can never occur, even if N and/or s are zero.

2. However, as long as m is not one, there are always some (large) values of N and s that make Stage 4 impossible.

3. If there is uncertainty about the value of m (i.e., the bank is unsure to what extent the Federal Reserve System will support it in a liquidity crisis), the "loaned-up" limit will probably occur at a value of L less than that calculated by using the expected value of m. How much less we are not equipped to say, on the basis of our too simple assumption that the bank never incurs any possibility of Stage 4. Scott suggests the possibility that each bank's "loaned-up" limit of L is based upon other banks', on the grounds that it is certain that the Federal Reserve System will not permit a general liquidity crisis to occur under any conditions. See p. 219 (especially footnote 15) of I. O. Scott, "The Changing Significance of Treasury Obligations in Commercial Bank Portfolios," Journal of Finance, Vol. XII, No. 2 (May 1957), pp. 213-222.

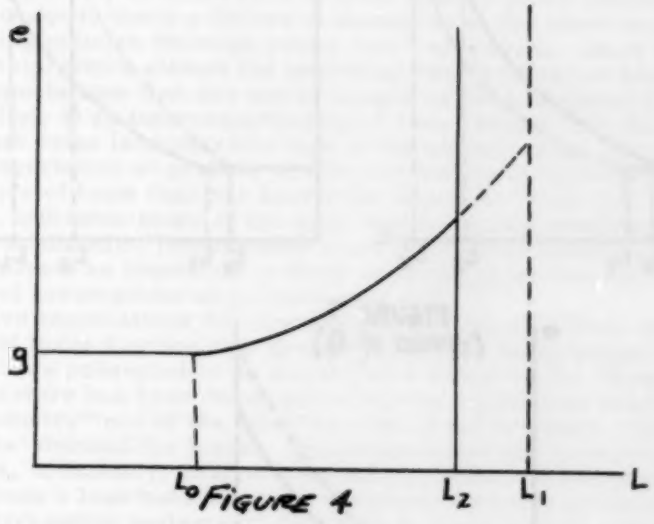


FIGURE 4

for the five parameters, s , k , q , g , and N , are illustrated in Figures 5 - 9 respectively; the solid line is the marginal cost for a lower value of the parameter and the dotted one for a higher, and L_0' and L_2' represent the values of L_0 and L_2 pertinent to the higher parameter value.

A fall in s and N and a rise in k , q , and g still have the same general effect of raising the marginal cost function, but the details of these shifts differ. Only s , k and N are capable of affecting the "loaned-up" limit (imposed by Stage 4 possibilities); and only g can affect the curve in its perfectly elastic range, although changes in s and k can alter the point at which the elasticity becomes finite. Although Figures 5 - 9 are illustrative and not empirically derived, they help to show how Federal Reserve System actions may achieve an impact upon bank portfolios; the traditional central bank policy weapons of reserve requirements, discount rate, and open-market operations are seen to operate primarily, in the world of the model, through the parameters, k , q , and g , respectively. And, if these (or any other) central bank measures should alter the bank's uncertainty about its deposit future, then there is an effect through s . A rise in g will generally inflict capital losses on the bank and, to the extent that these reduce N , will further raise the marginal cost of loans schedule of the bank.

The preceding discussion of the bank's supply of loans little resembles the current literature on credit "availability." Although a rise in bond interest rates is effective in shifting the schedule upward and/or to the left (through g , N , and possibly q or s), there is in the model nothing to indicate that small changes in interest rates will have large effects, a basic

FIGURE 5
(EFFECT of S)

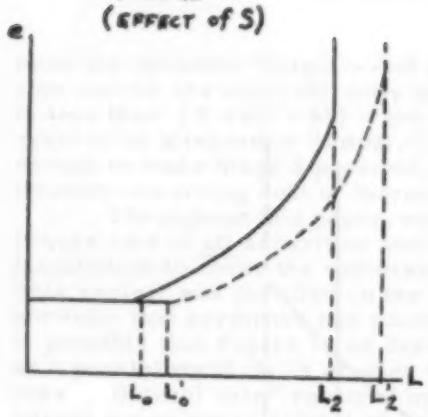


FIGURE 6
(EFFECT of K)

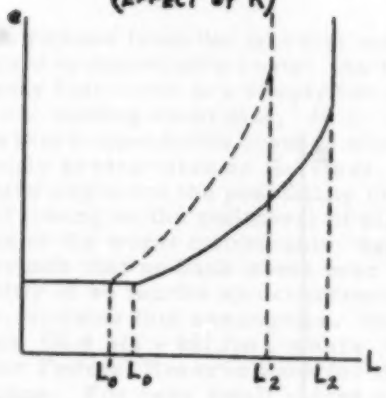


FIGURE 7
(EFFECT of Q)

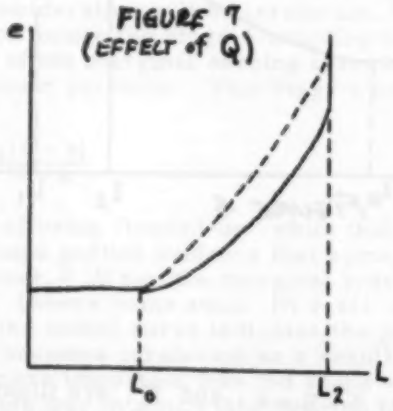


FIGURE 8
(EFFECT of G)

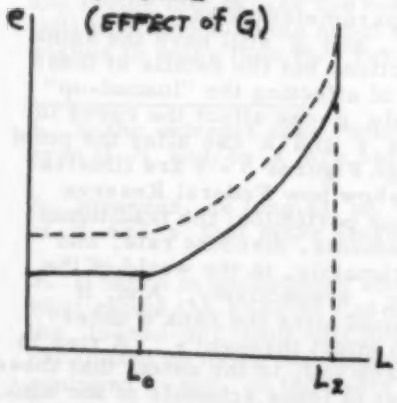
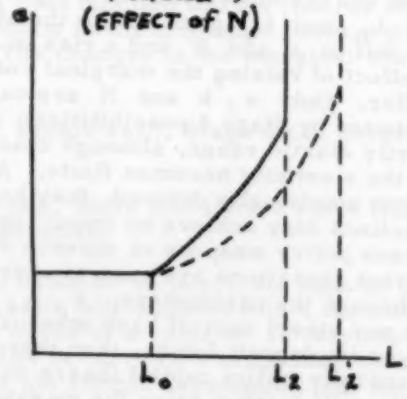


FIGURE 9
(EFFECT of N)



tenet of the "new theory of credit control."[1] Notice of this incompatibility is worthwhile, for it implies that availability theories rest upon assumptions different from, or supplementary to, those of the model. Some of the more obvious are: 1) explicit consideration of multi-period profit maximization, which might frequently mean a failure to maximize in the short-run; 2) oligopolistic pricing policies (through prime rate conventions, usury laws or market leadership) which distort the individual bank's effective loan demand schedule; 3) expectations that are not so simple as here assumed (i.e., unitary elasticity); 4) an inter-relationship of loan demand with the level of security interest rates (and possible lags in the adjustments); and 5) greater concern with uncertainty of profits, credit-worthiness of borrowers and the non-price factors of loans than has here been included. This list, certainly not exhaustive, indicates some of the ways that the model would have to be extended if the availability implications were to be more closely approached. It is perhaps at least as important to know what cannot be derived from a particular set of assumptions as to know what can.

To derive implications for monetary policy directly from the bank's marginal cost of loans function is a great temptation, but a dangerous one. For the model here presented is no more than a theory of the "firm;" no theoretical structure has been developed about the adjustment mechanism of the banking "industry" nor of the other sectors of the economy, from whence comes the banks' demand for loans. Knowledge about macro-economic behavior requires, ultimately, macro-economic analysis. The hypotheses about the individual bank's loan behavior here developed are only of value, from the viewpoint of aggregative analysis, if they help to place it upon a more sound micro-economic foundation.

7. Conclusion

Not too long ago, it would have been considered presumptuous to claim that knowledge of the bank's portfolio of earning assets could be useful to analysis of monetary theory or policy. Once economists had become convinced that commercial banks really could "create" money,[2] they became enamored with the fact that the amount of money thus created was limited by the quantity of currency and reserves which the central bank issued. To give precision to the formula relating currency and reserves to the money supply, all that was needed was knowledge of the public's and the banks' propensities to hold currency and the circumstances in which the banks keep excess reserves. Behind such total concern for the money supply always lies the assumption, explicit or implicit, that the velocity of money (or, in more acceptable modern terminology, the relationship of aggregate demand to the money supply) was constant, or at least fairly predictable.

In a world where money was used primarily for transactions purposes, and where only a small and relatively unchanging fraction of the total was used as a way of holding wealth, the quantity theory, at least in its more

---

1. This phrase is the title of an excellent summary of the innovations to monetary theory of the 1950s by A. Lindback, (Uppsala: 1959).

2. And this was not so long ago as we would like to think. As late as 1921, Professor Cannan wrote: "If cloak-room attendants managed to lend out three-quarters of the bags entrusted to them... we should certainly not accuse the cloak-room attendants of having 'created' the number of bags indicated by the excess of bags on deposit over bags in the cloak-rooms." Page 31 of "The Meaning of Bank Deposits," Economica, No. 1, January 1921, pp. 28-36.

sophisticated presentations, would be a good approximation of reality. And neglect of bank portfolios, beyond the problem of changes in currency and reserve holdings (either required or desired) would be thoroughly consistent with the theory. What the bank's earning assets were technically labelled--securities, advances, call loans, etc.--might, and obviously would, matter to the banker, but to the monetary theorist they would be just different ways of placing active money balances in the hands of the public.[1] Should the bank decide to somewhat alter its portfolio from securities to loans, the final result would be merely to reduce the spending potential of those individuals who increased their holdings of securities while increasing, to the same extent, the spending potential of the recipients of the new loans.

Recognition that "idle" balances are neither an insignificant nor an unchanging fraction of the total money supply and the fact that contemporary governments provide a wide range of default-free forms of wealth-holding, from currency to consols (varying each from the other only slightly, in liquidity and yield), forces upon monetary theorists an entirely different mode of analysis. While the theory of the determination of the velocity of money was dramatically revised by Keynes' General Theory, the relevance of this to monetary policy has only gradually become apparent. It is only during the current decade that monetary authorities have finally become as concerned with the manner in which the banking system makes money available to the public as with the total quantity. A movement by banks out of securities into loans cannot be uninteresting to monetary policy: the bond-buying public may be merely transferring a part of its wealth to a less liquid form, while the recipients of new loans are almost certainly increasing their spending.

Thus the division of the bank's earning assets between securities and loans is of relevance not only to the bank itself (concerned as it is with liquidity and profits) but also to analysis of the inflationary impact of different allocations of a given supply of money. Coincidentally--though perhaps providentally so--the bank's choice between greater liquidity and greater earnings is also society's choice between lesser and greater inflationary forces. Most of the weapons of contemporary monetary policy can be understood, and are in fact proposed, as an effort by the central bank to "encourage" (where that word covers a spectrum of meanings from "suggest" to "force") the commercial banks to hold assets which are relatively more liquid (though the means of achieving this often involves making such assets less liquid). It it toward an improved understanding of the ways in which various aspects of monetary policy affect the bank's choice between different earning assets, and hence aggregate demand, that this paper is aimed.

Appendix A: The "Deposit-Low"

Most banks have many depositors and the typical deposit transaction involves but a small fraction of the owner's account. For this reason, any attempt to derive theoretically the shape of the distribution of "deposit-lows" for a bank must recognize that any particular "deposit-low" is the result of a long series of individual deposit withdrawals and additions. Avoiding the question of what causes a depositor to alter the size of his account, we will deduce the "deposit-low" distribution on various assumptions about the prob-

1. Rather curiously, through most of the nineteenth century, a great many bankers and economists did believe that certain classes of bank assets were inherently less inflationary than others. But the "real bills" doctrine was founded not on a belief that the money thereby created was any less "active," but on a mistaken notion concerning aggregate supply.

ability of each deposit account transaction being an addition to or depletion of total deposits.

In line with the static nature of expectations in the text, let us assume that the bank "expects" no change in its total deposits during the ensuing period; this may be interpreted to mean that it believes that every dollar of transactions in its deposit accounts has a 50-50 chance of being a withdrawal of a dollar or a deposit of a dollar. It is certainly true, then, that the expected change of deposits, no matter how many transactions occur, will then be zero. But there will still be finite probabilities attached to "deposit-lows" less than zero. In general, if (2N) transactions occur, the probabilities of the "deposit-lows" are given by:

$$\Pr[0] = \frac{(2N)!}{2^{2N} \cdot (N!)^2}$$

$$\Pr[-2x] = \Pr[-2x+1] = \frac{(2N)!}{2^{2N} \cdot (N-x)! \cdot (N+x)!}$$

where x is a positive integer, $\Pr[-2x]$ means the probability of a "deposit-low" of $(2x)$ dollars less than initial deposits, and $1 \leq x \leq N$. [1] A smoothed picture of such a discrete probability function is given in Figure A-1, [2] where

---

1. Cf. J. L. Doob, Stochastic Processes, (New York: 1953), pp. 106-108. Let L be the "deposit-low" and A the level of deposits after the completion of all $(2N)$ transactions. Doob proves, with the "reflection principle," that

$$\text{Prob}\,[L < -2x] = 2\,\text{Prob}\,[A \leq -2x] - \text{Prob}\,[A = -2x]$$

Since

$$\text{Prob}\,[L = -2x] = \text{Prob}\,[L < -2x+1] - \text{Prob}\,[L < -2x]$$

it follows that

$$\text{Prob}\,[L = -2x] = \text{Prob}\,[A = -2x+1] + \text{Prob}\,[A = -2x]$$

But $\text{Prob}\,[A = -2x+1] = 0$ (since an even number of $+\$1$ or $-\$1$ transactions can never result in an odd number of dollars withdrawn). Thus,

$$\text{Prob}\,[L = -2x] = \text{Prob}\,[A = -2x]$$

It can easily be shown also that

$$\text{Prob}\,[L = -2x] = \text{Prob}\,[L = -2x+1]$$

Thus the distribution of "deposit-lows" is the same (with twice the density) as the distribution of end-of-period deposits over the range of net decreases of deposits. Clearly, as N becomes larger, the "deposit-low" distribution, being binomial, approaches a truncated (at zero) normal distribution.

2. It should perhaps be noted that this distribution differs from Patinkin's receipts-expenditures distribution for individuals because there is here no assumption that withdrawals must equal accretions (i.e., both equal to N). Cf. D. Patinkin, Money, Interest and Prices, Evanston: 1956, Chapter 7

the solid curve represents the distribution for (2N) transactions and the dotted curve for a larger number of total transactions.[1]

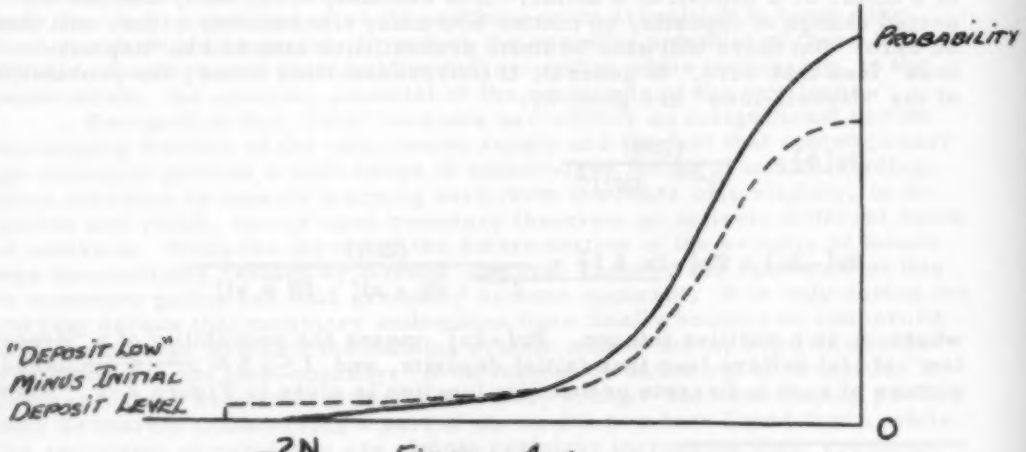


FIGURE A-1

It is unlikely that any bank is in a position, however, to consider each transaction as totally independent of all or any previous transactions in the deposit accounts. At one extreme, if there is only one bank in the economy and there is no change in the public's desire to hold cash (as opposed to deposits), then any withdrawal must appear later as an addition; there is definitely an inverse relation between the probability that a given transaction will be an addition and the proportion of previous transactions which were additions. The more usual case, however, is that of a bank which experiences or expects, or fears, positive correlation between the probability that a given transaction will be a deposit and the relative number of previous transactions which were deposits.

As an example of this, suppose that, whenever a majority of the previous transactions have been deposits, the probability that the next transaction will also be a deposit is somewhere between .5 and unity. In order to maintain the assumption of static expectations, the probability of a withdrawal when a majority of the previous transactions were withdrawals must be equal to the probability of a deposit when a majority were deposits. The earlier

---

and Appendix to Chapter 7 (by Aryeh Dvoretzky). Note especially the difference between Figure A-1 below and Patinkin's Figure 9, p. 92.

1. The slope of the frequency distribution is positive over its entire range, from (-2n) to zero. It increases first at an increasing rate, but at a de-

formulae are now seen as the special case where these probabilities are .5. A rough picture of the effect of raising them above .5 is shown in Figure A-2, where the solid curve represents the same distribution as in Figure A-1, and the dotted curve the distribution when autocorrelation is introduced in the above manner. It is doubtful if the autocorrelation of deposit transactions is sufficiently large, for most banks, to cause a negative slope in the frequency distribution of "deposit-lows," but it may suffice to keep the probability of relatively low "deposit-lows" from being so small as to be negligible.

The "triangular" distribution has been chosen in the text as the closest simple approximation to this distribution. If a straight line were fitted through the distributions of Figure A-2, it could be seen that the "triangular" distribution understates the probability of occurrence of extremely low and zero-neighborhood "deposit-lows" and overstates the probability of the middle range of "deposit-lows."

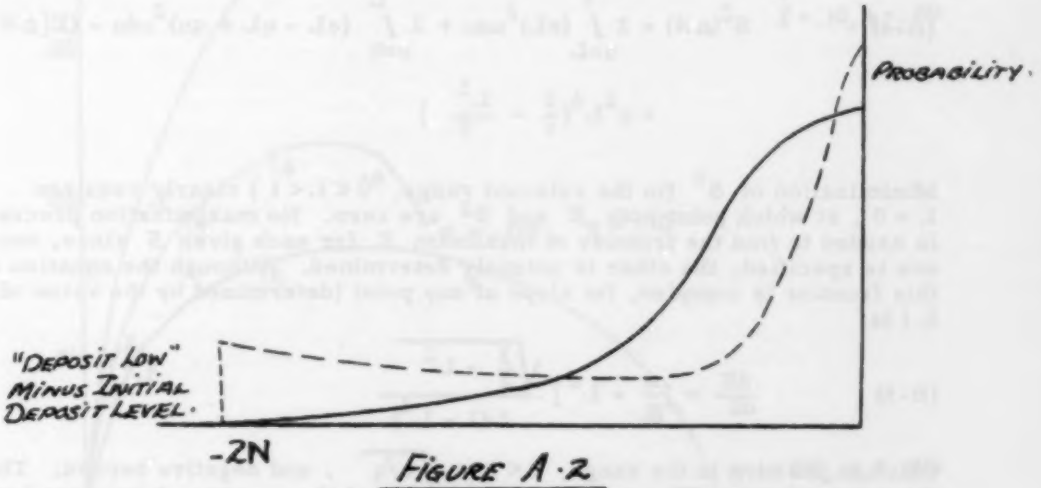


FIGURE A-2

Appendix B: Variance of Profits

Because the consideration of variance of earnings involves great complication of the model, only a very simplified version of it will be discussed here. In addition to the assumptions of the text, it is assumed that the bank has no cash requirements ($k = 0$), it has no net worth ($N = 0$), there is no "excess-collateral rate" ($m = 0$), securities have no earnings ($g = 0$) and no possibility of price change ($a = 0$), and it is considered possible that all deposits be withdrawn during the period ($s = 0$). Thus the accounting identity (1) of the text becomes:

(B-1) $\qquad 1 = B + L$

---

creasing rate in the area to the immediate left of zero. The inflexion point occurs at $x^2 = \frac{1}{2}(N + 1)$.

The distribution of "deposit-lows" is: $f(u) = 2u$; and the expected additions to net worth function is simple:

$$(B\text{-}2) \qquad E[\Delta N] = 2\int_{u=L}^{1}(eL)udu + 2\int_{u=0}^{L}(eL - qL + qu)udu = eL - \frac{qL^3}{3}$$

Maximization of expected profit implies that:

$$(B\text{-}3) \qquad L = \sqrt{e/q}$$

A regular maximum will occur, with L between zero and one, as long as $e < q$ (both are positive).

The variance of profit, written $S^2(\Delta N)$, is:

$$(B\text{-}4) \qquad S^2(\Delta N) = 2\int_{u=L}^{1}(eL)^2udu + 2\int_{u=0}^{L}(eL - qL + qu)^2udu - (E[\Delta N])^2$$

$$= q^2L^4\left[\frac{1}{6} - \frac{L^2}{9}\right]$$

Minimization of S^2 (in the relevant range, $0 < L < 1$) clearly requires $L = 0$, at which point both E and S^2 are zero. No maximization process is needed to find the frontier of maximum E for each given S since, once one is specified, the other is uniquely determined. Although the equation of this frontier is complex, its slope at any point (determined by the value of L) is:

$$(B\text{-}5) \qquad \frac{dE}{dS} = \left[\frac{e}{q} - L^2\right]\frac{\sqrt{\frac{3}{2} - L^2}}{L(1 - L^2)}$$

which is positive in the range $0 < L < \sqrt{e/q}$, and negative beyond. Thus, the expected gain (E) can be increased only at the expense of increased variance of gain (S^2) up to the point of maximum E . These opportunity loci are plotted in Figure B-1 for three sets of values of e and q .

On the assumption that bankers are "risk-averters" and "diversifiers,"[1] it can be easily seen from Figure B-1 that they will choose L at least somewhat smaller than that value which maximizes E . However, the change in the portfolio induced by a change in e or q can only be guessed unless the bank's complete preference function between E and S is specified, a task which will not be attempted.

Three conclusions may be drawn from this discussion:

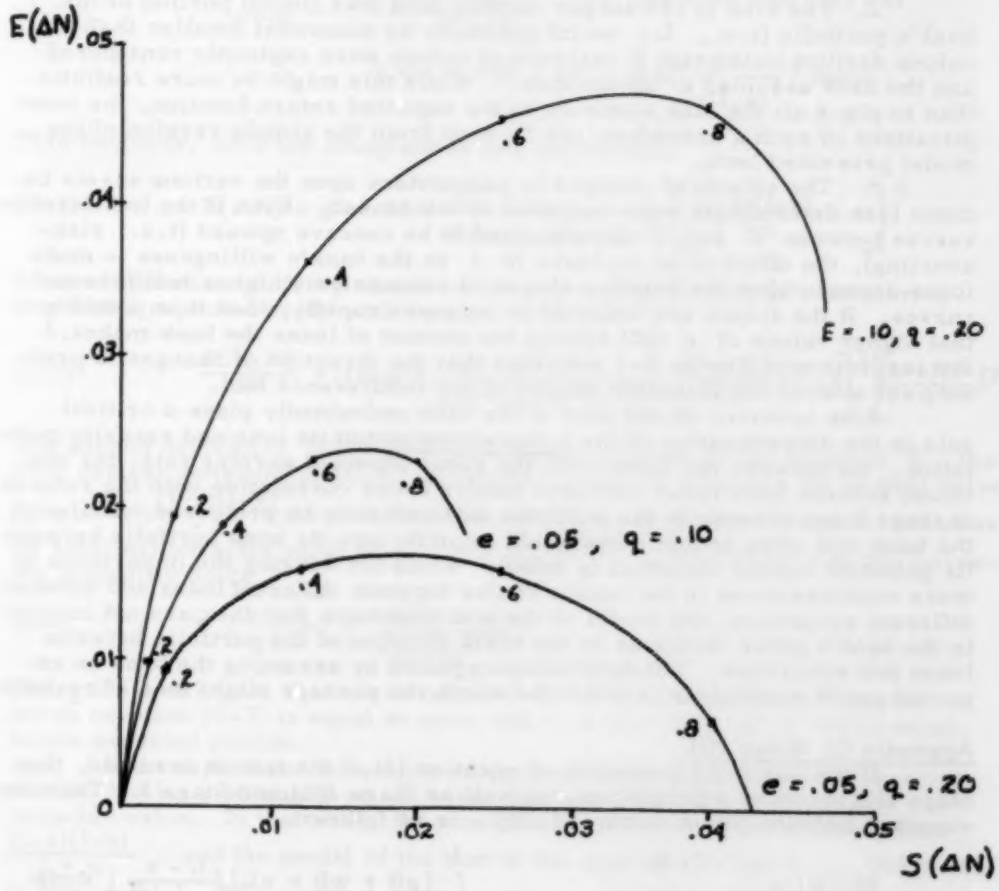
1. The assumption that the bank maximizes E and neglects S entirely is equivalent in the Tobin sense, to assuming that the bank is on the border between risk-averting and risk-loving.[2] This is not as serious as the

---

1. Cf. Tobin, op. cit., pp. 16-22. The labels "risk-averter" and "diversifier" are applied to those investors whose indifference curves between E and S are concave upward. For opportunity loci such as those of Figure B-1, "plungers" and "risk-lovers" might also diversify (i. e., choose L not equal to zero or one), but "diversifiers" necessarily will diversify.

2. Ibid., p. 19. In Tobin's equation (3. 7) maximizing expected return im-

FIGURE B-1



(THE NUMBERS BESIDE THE POINTS ARE THE FRACTION OF TOTAL ASSETS HELD IN LOANS AT THAT POINT; THE CURVES END AT THE POINT, L = 1, EXCEPT FOR THE CURVE, e = .05, q = .20, WHICH ENDS AT L = .86 WHERE E BECOMES NEGATIVE.)

words imply since much of the bank's risk is reflected in the model of the text in the expected profit function itself.

2. The size of the larger earning (and less liquid) portion of the bank's portfolio (i.e., L) would generally be somewhat smaller than the values derived in the text if variance of return were explicitly considered and the bank assumed a "diversifier." While this might be more realistic than to place all the risk elements in the expected return function, the complications of such a procedure can be seen from the simple version of the model presented here.

3. The effects of changes in parameters upon the various assets become less determinate when variance is introduced. Even if the indifference curves between E and S are assumed to be concave upward (i.e., risk-averting), the effect of an increase in e on the bank's willingness to make loans depends upon the relative slopes of successively higher indifference curves. If the slopes are believed to increase rapidly, then it is possible that higher values of e will reduce the amount of loans the bank makes.[1] But inspection of Figure B-1 indicates that the direction of changes is probably not altered for plausible shapes of the indifference loci.

Risk aversion on the part of the bank undoubtedly plays a critical role in the determination of the composition within its loan and security portfolios. As between two loans with the same expected earning rate, the one whose returns have lower variance and/or lower correlation with the returns of those loans already in the portfolio will certainly be preferred; similarly, the bank will often accept a low-yield security into its bond portfolio because its potential capital variation is small. While not denying the importance of these considerations in the bank's choice between different loans and between different securities, the model of the text maintains that they are not crucial in the bank's prior choice as to the basic division of the portfolio between loans and securities. The determinancy gained by assuming the bank an expected profit maximizer is felt to be worth the perhaps slight loss of realism.

Appendix C: Stage 2(i)

If the left-hand inequality of equation (5) of the text is assumed, then Stage 2(i) becomes a possibility, as well as Stage 2(ii) and Stage 3. Then the expected addition to net worth $E[\Delta N]$, is as follows:

$$E[\Delta N] = \int_{w=-1+\frac{(1-s)(1-k)}{B}}^{a} \int_{u=s}^{1} [gB + wB + eL]\left[\frac{u-s}{a(1-s)^2}\right] dudw$$

$$+ \int_{w=-a}^{-1+\frac{(1-s)(1-k)}{B}} \int_{u=1-\frac{(1+w)B}{1-k}}^{1} [gB + wB + eL]\left[\frac{u-s}{a(1-s)^2}\right] dudw \qquad \text{(C-1)}$$

---

plies a marginal utility of return (U'(R)) which is constant with respect to changes in R.

1. This possibility has been called an "income" effect by J. Aschheim in "Open-market Operations versus Reserve-requirement Variation," Economic Journal, December 1959, pp. 697-704, where it is suggested that greater bank profits, ceteris paribus, will reduce the quantity of loans.

$$+ \int_{w=-a}^{-1+\frac{(1-s)(1-k)}{B}} \int_{u=s}^{1-\frac{(1+w)B}{1-k}} [gB + wB + eL - q(1-k) + q(1+w)B + q(1-k)u][\frac{u-s}{a(1-s)^2}]\, dudw$$

This becomes, once the integrations are performed,

$$(C\text{-}2) \qquad E[\triangle N] = gB + eL - \frac{q(1-a)^4}{24Ba(1-s)^2(1-k)^2}\left[\frac{(1-s)(1-k)}{1-a} - B\right]^4$$

After elimination of L by means of the accounting identity (1), the derivative of the expected profit with respect to B is:

$$(C\text{-}3) \qquad \frac{dE[\triangle N]}{dB} = -(e-g) + \frac{q(1-a)^4}{24B^2a(1-s)^2(1-k)^2}\left[\frac{(1-s)(1-k)}{1-a} - B\right]^3\left[\frac{(1-s)(1-k)}{1-a} + 3B\right]$$

The second derivative (d^2E/dB^2) is negative throughout the range of B, the first derivative (equation (C-3)) is negative when B assumes its largest value (i.e., $B = \frac{(1-s)(1-k)}{1-a}$), and the first derivative is positive when B assumes its smallest possible value (i.e., $B = \frac{(1-s)(1-k)}{1+a}$), provided that the following inequality holds:

$$(C\text{-}4) \qquad 0 < \frac{e-q}{q} < \frac{2a^2(2-a)}{3(1+a)^2}$$

Thus, if (C-4) holds, there is a value of B in the permissible range for which equation (C-3) is equal to zero; and it is that value of B which maximizes expected profits.

If the right-hand inequality of equation (C-4) does not hold, expected profits decline continually as B rises from its smallest to its largest permissible value. In that case, the maximizing value of B is smaller than $\frac{(1-s)(1-k)}{1+a}$, and the model of the text is the appropriate one (i.e., Stage 2(i) is impossible). The common sense argument for choosing the assumption of the text that there is no Stage 2(i) is greatly strengthened by a consideration of the parameter values required if the right-hand inequality of (C-4) is to hold: the table below gives the highest possible values of (e - g), consistent with the existence of a Stage 2(i), for several combinations of values of a and q:

TABLE C-1

	q = .05	q = .10	q = .30
a = 0	0	0	0
a = .1	.0005	.0011	.0033
a = .25	.0023	.0047	.0141
a = .50	.0055	.0111	.0333

Inasmuch as e generally could be expected to exceed g (on any definition of loans and securities) by at least one or two per cent, surprisingly--if not implausibly--large values of a and q are required if the bank is to choose to hold enough securities to be able to meet its worst deposit depletions by securities sales along--even if bond prices rise to their highest conceivable value.

Nevertheless, the implications for the bank's optimum asset portfolio of changes in various parameters can be determined by taking partial derivatives of (C-3) while holding that equation equal to zero. The changes in optimum B are as follows:[1]

$$\partial B/\partial e < 0$$
$$\partial B/\partial g > 0$$
$$\partial B/\partial q > 0$$
$$\partial B/\partial s < 0$$
$$\partial B/\partial k < 0$$

The sign of $\partial B/\partial a$ cannot be determined generally. That these signs are the same as those derived in the text, on the assumption that Stage 2(i) does not exist, is partial evidence that the influence of parameter changes upon optimum portfolios does not depend critically upon this choice of assumption.

Appendix D: Time Deposits

No distinction has been made, anywhere in the text, between different types of deposits, a neglect which will be here repaired by introducing time, as well as demand, deposits. The most obvious difference between the two classes of deposits lies in their cost, for time deposits are still permitted to, and do, earn interest for the depositor. On the other hand, the "earning power" of time deposits may exceed that of demand deposits in either or both of two ways. First, it is an institutional fact in most countries that the cash, or low-yield, reserve requirements on time deposits are lower than those on demand deposits. And second, to the extent that the lowest conceivable "deposit-low" is raised by the addition of less volatile time deposits, the bank may hold a larger portion of its portfolio in less liquid, but higher earning, assets.[2]

Banks attract new time deposits by offering to pay higher interest rates on them. Such increases in time deposits may "come" from several places: by a transfer from 1) the bank's own demand deposits; 2) the public's deposits with other banks; and 3) deposits and shares of the non-bank financial intermediaries.[3] From the point of view of the individual bank, the relevant division is between the first case and the second and third cases.[4]

---

1. The arithmetic is sufficiently messy that only the first derivatives have been calculated.

2. Throughout this section, we neglect the possibility that greater time deposits may imply lesser private (or other financial intermediary) lending, and hence higher lending rates for the bank. Cf. J. Tobin, "Financial Intermediaries and the Effectiveness of Monetary Controls," Cowles Foundation Discussion Paper No. 63 (January 1959).

3. It is assumed that the public's currency needs are fixed.

4. From the point of view of the banking system, the most interesting di-

Here it will be assumed that the bank cannot, through its own volition, alter the total quantity of its deposits, but only, by changing its time deposit interest rate, induce some of its depositors to hold time instead of demand deposits. The fraction of total deposits held as time deposits is assumed to grow from zero, as the time deposit interest rate (i) increases from zero, to some maximum value at which point further increases in i cannot induce further shifts from demand to time deposits. Thus, time deposits as a fraction of total deposits (T) will lie between zero and some maximum level (T*) depending upon i , where T* is less than one.

This transfer of funds from demand to time deposits will affect the values of k and s . The overall cash requirements, as a fraction of total deposits, will be:

$$k = k_d - (k_d - k_t)T \tag{D-1}$$

where requirements on demand deposits (k_d) are assumed to exceed those on time deposits (k_t) . The value of s , and the frequency distribution of the possible "deposit-lows," is not so easily calculated. It will depend upon the lowest possible "deposit-low" of time deposits (as a fraction of initial time deposits), s_t , and that of demand deposits (as a fraction of initial demand deposits), s_d ; but it will also depend upon the form of the distribution of each of these "deposit-lows" and the covariance of the changes of each of these two types of deposits. If the frequency distribution of each type's "deposit-low" is assumed "triangular" (with a lowest possible "deposit-low" of s_t and s_d for time and demand deposits, respectively) and the two changes are uncorrelated, the resulting distribution of "deposit-lows" for the total of deposits will not be "triangular" but will have a convex (from below) segment over very low "deposit-lows" and a concave segment for "deposit-lows" near unity.[1] Since the assumption of zero covariance between

---

vision is between the first two cases and the third case.

1. If the fraction in time deposits is small enough that

$$T \leq \frac{1 - s_d}{(1-s_d)+(1-s_t)} \tag{D-2}$$

the frequency distribution of "deposit-lows" of total deposits, f(u) , will be:

$$f(u) = \frac{2}{3}\,\frac{[u-Ts_t-(1-T)s_d]^3}{T^2(1-T)^2(1-s_d)^2(1-s_t)^2} \quad \text{where: } Ts_t+(1-T)s_d < u < T+(1-T)s_d$$

$$= \frac{2}{3}\,\frac{3[u-Ts_t-(1-T)s_d]^2 - 2T(1-s_t)}{(1-T)^2(1-s_d)^2} \quad \text{where: } T+(1-T)s_d < u < (1-T)+Ts_t$$

$$= 2[u-Ts_t-(1-T)s_d]\left[\frac{1}{(1-T)^2(1-s_d)^2} + \frac{1}{T^2(1-s_t)^2}\right] - \frac{4(1-T)(1-s_d)}{3T^2(1-s_t)^2}$$

$$- \frac{4T(1-s_t)}{3(1-T)^2(1-s_d)^2} - \frac{2[u-Ts_t-(1-T)s_d]^3}{3T^2(1-T)^2(1-s_d)^2(1-s_t)^2} \quad \text{where: } (1-T)+Ts_t < u < 1$$

"deposit-lows" of time and demand deposits leads to an understatement of frequencies in the low "deposit-low" segment, and an overstatement in the high range, it will be convenient, and probably less unrealistic, to assume that the aggregate "deposit-low" distribution is also "triangular" with the point of zero probability, s, occurring at:[1]

$$(D\text{-}4) \qquad s = Ts_t + (1-T)s_d$$

The costs of time deposits, iT, are subtracted from equation (8) of the text to give the expected addition to net worth when the bank considers a time deposit policy:

$$(D\text{-}5) \qquad E[\Delta N] = gB + e(1+N-k-B) - \frac{q}{3x^2}(x-B)^3 - iT$$

where, for simplicity, bond prices are assumed not to fluctuate (i.e., a equal to zero) and, for brevity, x is written for $(1-s)(1-k)$. This expression, (D-5), is maximized by the bank not only with respect to B, which yields equation (12) of the text (with a equal to zero), but also with respect to i, for the interest rate paid on time deposits is now a variable under the control of the bank. Differentiation of (D-5) with respect to i yields, after substitution for the profit-maximizing value of B by means of equation (12):

$$(D\text{-}6) \qquad \frac{\partial E[\Delta N]}{\partial i} = e(k_d-k_t)\frac{dT}{di} - (e-g)\left[1 - \frac{2}{3}\sqrt{\frac{e-g}{q}}\right]\frac{dx}{dT}\cdot\frac{dT}{di} - \frac{d(iT)}{di}$$

where

$$(D\text{-}7) \qquad \frac{dx}{dT} = \frac{d(1-s)(1-k)}{dT} = (1-s_d)(k_d-k_t) - (1-k_d)(s_t-s_d) - 2T(s_t-s_d)(k_d-k_t)$$

Clearly, if i becomes sufficiently large that dT/di is zero (i.e., time deposits are at their maximum level, T^*), equation (D-6) will be negative; for some high values of i, expected profits will be reduced if time deposit rates are raised. At the other extreme, when i, and hence T, is zero, equation (D-6) is positive; costs are not increased as much as expected profits when i is increased slightly from zero. Thus, expected profits at first increase and later decrease as i, and hence T, is raised. We can conclude, on the above assumptions, that the optimum proportion of total deposits in time deposit accounts will always be greater than zero.[2] Moreover

---

The distribution, $f(u)$, will be "triangular" only if s_t is one, i.e., time deposits never decline. If the direction of the inequality of (D-2) is reversed, the exact form of the distribution is altered, but the general shape is the same (and the distribution is "triangular" only if s_d is one).

1. This value of s is the lowest conceivable "deposit-low" of the distribution (D-3).

2. W. L. Smith uses a simplified version of equation (D-6), on the assumption that e equals g (i.e., neglecting the liquidity differences between assets), to show that it is profitable for the bank to induce switches from demand to time deposits, by raising time deposit rate, if:

$$e > \frac{i}{k_d - k_t}\left[1 + \frac{1}{\eta}\right]$$

this conclusion still follows if time deposits are assumed to be reserve-saving though not deposit-stabilizing (i.e., $k_d > k_t$, $s_d = s_t$), or vice versa (i.e., $k_d = k_t$, $s_d < s_t$). Of course, there are many reasons beyond the scope of the model why the optimum amount of time deposits might be zero for a particular bank. Most obvious of these are that administrative considerations place a minimum on the amount of such deposits which the bank will wish to attract and/or that a time deposit rate markedly above zero is required to induce any time deposits at all. On the other hand, there are reasons other than expected profits (as defined here) for desiring such relatively stable deposits even if not profitable--for example, concern by the bank for the variability of its expected profits (see Appendix B).

Appendix E: Uniform Distribution of "Deposit-Lows"

In order to indicate that the results of the model of the text do not depend too critically upon the type of distribution assumed for the bank's "deposit-lows," the same model is here worked through briefly on the alternative assumption that the "deposit-lows" are uniformly distributed over the range from s to unity; that is,

$$(E\text{-}1) \qquad f(u) = 1/(1-s)$$

The expected addition to net worth is then exactly as given by equation (7) of the text with the appropriate alteration in the distribution of u. The analog to equation (8), expected profit after the integrations have been performed, is:

$$(E\text{-}2) \qquad E[\Delta N] = gB + eL - \frac{q}{2(1-s)(1-k)}\,[(1-s)(1-k) - B]^2 - \frac{a^2 q B^2}{6(1-s)(1-k)}$$

And the optimum proportion of securities (see equation (12)) is:

$$(E\text{-}3) \qquad B = \frac{(1-s)(1-k)}{(1 + a^2/3)}\,\left[1 - \frac{e-g}{q}\right]$$

It should be noted that the parameters enter in a less complex, but similar, manner (see pages 22-23).

The first and second partial derivatives of B have the following signs:

TABLE E-1

		$\partial B/\partial x$	$\partial^2 B/\partial x \partial y$ where					
			y = a	y = q	y = g	y = e	y = k	y = s
where	x = s	-	+	-	-	+	+	0
	x = k	-	+	-	-	+	0	
	x = e	-	+	+	0	0		
	x = g	+	-	-	0			
	x = q	+	-	-				
	x = a	-	-					

where η is the time deposit interest elasticity of time deposits. Of course, the above conclusions follow in this special case as well. Cf. footnote 7, p. 544 of W. L. Smith, "Financial Intermediaries and Monetary Controls," Quarterly Journal of Economics, Vol. 73, No. 4 (November 1959), pp. 533-553.

Again, there is no basic difference between this table and that on page 335 although here more second partials are zero and there is no possibility of different signs in different regions of the parameters.

Appendix F: End-of-Period Deposits

In the text it is assumed that the "most radical" adjustment of the bank's portfolio is required at the moment when its deposits reach their lowest point of the period. If the bank is able to postpone this adjustment until the end of the period by means of short-term borrowing, then the relevant distribution for its profit calculations is that of end-of-period deposits. It is interesting to compare the implications for the bank's portfolio of this alternative assumption about the relevant deposit distribution.

In Appendix A it is shown, under certain assumptions, that the "deposit-low" distribution is simply a truncated (at the level of initial deposits) duplication (of twice the density) of the end-of-period deposit distribution. Since this "deposit-low" distribution is approximated in the text by a "triangular" distribution,

$$f(u) = \frac{2(u-s)}{(1-s)^2} \quad \text{for} \quad 0 < s < u < 1 ,$$

comparison between the two assumptions is facilitated if the distribution of end-of-period deposits, $f^*(v)$, is approximated by an "isosceles" distribution:

$$f^*(v) = \frac{(v-s)}{(1-s)^2} \quad \text{for} \quad 0 < s < v < 1$$

$$f^*(v) = \frac{(2-s-v)}{(1-s)^2} \quad \text{for} \quad 1 < v < (2-s)$$

The distribution of v for values greater than unity need not detain us, for the possible earnings obtained from net increases in deposits are quite independent of the start-of-period portfolio decision.

Therefore the expected addition to net worth may be written, analogously to equation (7) of the text, as:

$$(F\text{-}1) \qquad E[\Delta N] = \int_{w=-a}^{a} \int_{v=s}^{1} [gB + wB + eL] \left[\frac{v - s}{2a(1-s)^2}\right] dv\, dw$$

$$+ \int_{w=-a}^{a} \int_{v=s}^{1-\frac{(1+w)B}{1-k}} [-q(1-k)+(q(1+w)B+q(1-k)v] \left[\frac{v - s}{2a(1-s)^2}\right] dv\, dw$$

$$+ \frac{1}{2} \int_{w=-a}^{a} [gB + wB + eL] \left[\frac{1}{2a}\right] dw$$

Maximization yields a value for B , the optimum securities holding, analogous to equation (12) of the text:

$$(F\text{-}2) \qquad B = \frac{(1-s)(1-k)(1+\frac{a^2}{3})}{1 + a^2} \left\{ 1 - \sqrt{1 - \frac{(1+a^2)[1-2(\frac{e-g}{q})]}{(1 + a^2/3)^2}} \right\}$$

The only difference between equation (F-2) and equation (12) is that the term, $(e-g)/q$, is multiplied by a factor of two in equation (F-2). For equal values of all parameters, the present formulation of the deposit distribution therefore implies a lower amount of securities and a larger amount of loans.[1] The directions of change in the optimum portfolio induced by variations in the parameters are, however, not affected by this concern for end-of-period deposits instead of the "deposit-low." Of course, the results would be different for other specific forms of the distributions, but this exercise indicates that the conclusions of the model are not critically dependent upon the deposit distribution assumption.

1. A corner maximum with no securities held in the optimum portfolio, now occurs when $(e-g)/q$ is greater than .5, whereas, on the assumption of the text, the corner maximum occurred when $(e-g)/q$ was greater than 1.0.

Summary: Behavior of profit margins in the United States economy is empirically investigated in its various aspects. First, an aggregative examination indicates distinct secular and cyclical responses of the percentage gross margin in relation to the percentage overhead cost in the manufacturing sector. A similar comparison among a number of selected sectors of the economy seems to favor a hypothesis that their differential patterns of price-cost behavior are for the most part governed by differences in supply and demand conditions without making any reference to industrial concentration. A comparison of manufacturing firms making profits and losses points to limitations of demand as the most influential factor in determining differential price-cost relationships among them. Next, persistent interindustry differentials in percentage net profit margins among manufacturing industries are found to be explained more or less by differences in capital intensity, industrial concentration, and growth and fluctuations in production. However, an examination of cyclical patterns of price-cost behavior among various manufacturing industries leads to a rejection of the hypothesis that they are to be directly related to differences in their market structure or monopoly power. Finally, interindustry wage variations over time are statistically examined to assess interactions between cost elements and the price-cost structure.

This essay was submitted to Yale University in 1960 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics, Institute of Social and Economic Research, Osaka University.

Dissertation Committee: Professors Richard Ruggles (Chairman), William J. Fellner, Alan S. Manne.

*

I am greatly indebted to Professor Richard Ruggles who gave me his time generously. I am also grateful to Professor William Fellner, who read the manuscript. My thanks are due to the Cowles Foundation for Research in Economics at Yale University for part of the financial assistance. The basic data processing for part of the data used in this study was conducted at the MIT Computation Center while the author held an IBM research assistantship there. For generous assistance in data collection, I owe gratitude to the staff of the Industry Division of the Bureau of the Census, the Statistics Division of the Internal Revenue Service, and the Division of Productivity and Technological Developments of the Bureau of Labor Statistics.

K.S.

PRICE-COST STRUCTURE AND BEHAVIOR OF PROFIT MARGINS

Kazuo Sato

1. Introduction

Price theory or value theory is one branch of economics in which empirical verification of theory has lagged far behind theoretical development. Significant contributions of empirical nature within the analytical framework of theory have been rather scarce in this field. We need more intensive as well as extensive empirical studies to obtain a better understanding of the working of our price system. If we compare the present status of price theory with that of aggregative theory of employment and income, we find the contrast quite striking. The latter has attained remarkable progress in empirical applications both in terms of quality and quantity of academic output in the past two decades. Backwardness of value theory in its empirical development is certainly a serious drawback in the way of combining theories of value and employment successfully. We are still not very far from where we were a dozen years ago when Professor Fellner commented that "the process of including problems of the cost-price structure in the theory of employment is still in its early stages."[1]

There are many avenues of approach to the empirical realm of price theory. Case studies of individual firms and industries in pricing practices are one such. Studies of price flexibility are another. It is important, however, to have a broad, overall picture of the working of the price system before undertaking more detailed individual studies.

In this connection, our empirical concept which is most directly related to the theoretical one is perhaps that of profit margins. Cross profit margins as a percentage of sales can be linked to the concept of equilibrium price-cost relationships of the firm. Importance of profit margins in the context of price theory is well recognized in economic literature. To cite a few examples, gross profit margins, primarily determined by the degree of monopoly power, are the determinant of income distribution; or their short-term stability is a measure of price flexibility; or full-cost pricing is the actual pricing policy which sets price with a predetermined percentage margin added to average unit prime cost.

However, we still lack sufficiently conclusive empirical findings on behavior of profit margins. The present study attempts to help fill this gap in our knowledge through an empirical analysis of various aspects of profit margins in the United States economy. It tries to answer such questions as how percentage profit margins of different sectors behave secularly as well as cyclically, how they differ among successful and unsuccessful firms, whether monopoly power is an important factor in determining inter-industry differentials in profit margins as well as their cyclical behavior, and so on. Because of the availability of better data and clarity in concepts, the analysis is mostly confined to the manufacturing sector.

Various hypotheses on cyclical and secular patterns of behavior of percentage profit margins, presented in a number of previous studies, have been tested and the implications of them have been drawn. In Section 2, we examine secular and cyclical behavior of the percentage gross margin in relation to percentage overhead cost; it is found that in manufacturing as a whole for the period 1917-1956, the percentage gross margin is largely related to the secular level of their percentage overheads as well as the income tax schedule, but shows quite distinct, even though minor, cyclical responses

---

1. W. Fellner, "Employment Theory and Business Cycles," in H. S. Ellis, ed., A Survey of Contemporary Economics, Vol. I (Homewood, Illinois, R. D. Irwin, Inc., 1949), p. 86.

to changes in demand. A similar comparison among different sectors of the economy supports an hypothesis that differential patterns of price-cost behavior among them are for the most part governed by differences in the production structure with differential adaptations to fluctuations and to trends in demand. It is not necessary to make any reference to industrial concentration in this respect. Within the manufacturing sector a comparison of firms making profits and losses yields the conclusion that limitations of demand are the most influential factor in determining differential price-cost relationships among them. In Section 3, we go into less aggregated levels and undertake an interindustry study of manufacturing for the post-war period. Interindustry differentials in percentage net margins are found to be explained partly by differences in capital/sales ratios, industrial concentration, secular trends in demand and cyclical fluctuations in output. In particular, the importance of differential rates of industrial expansion in determining the secular profitability of industries is noted. An examination of differences in cyclical price-cost behavior among manufacturing industries leads to a tentative rejection of an hypothesis that the degree of market imperfection or monopoly power determines cyclical patterns of price-cost behavior among manufacturing industries. Rather it tends to suggest other factors--in particular, cyclical flexibility of demand. In the appendix where we try to investigate the relationships of the interindustry profit and wage differentials, we make a statistical test on determinants of variations in the interindustry wage structure; it rejects the hypothesis that differential changes in labor productivity is a dynamic factor in interindustry wage variations over time, but suggests differential rates of expansion, profitability, and industrial concentration as important factors in this connection.

2. Aggregative Study

A. Secular and Cyclical Behavior of Percentage Profit Margins, All Manufacturing, 1917-1956

1. Problem. The importance of percentage gross profit margins, i.e., the ratio of price minus average unit direct cost to price, for various economic problems such as measurement of monopoly power and price flexibility has been well recognized. Nevertheless we have had no conclusive empirical findings about how they would behave over business cycles and over the long run. As behavior of price-cost relationships is important not only for theory but also for policy formulation, it is worth while making a more intensive investigation of this subject.

2. Previous Study. Let us first review previous findings in this respect. The theory of imperfect competition suggested that the gap between price and marginal cost in short-run equilibrium may be considered as a measure of the degree of monopoly power.[1] If a firm maximizes its short-run profits, its price minus marginal cost (which is equal to marginal revenue) as a percentage of price at the equilibrium output, i.e., the degree of monopoly, is equal to the inverse of the elasticity of demand. Insofar as short-run profits are maximized, it follows that the price-marginal cost relationship is influenced neither by shifts in demand nor by changes in cost. It changed only if demand elasticity is changes at equilibrium output. The percentage gross margin, then, is determined by the elasticity of demand and the average cost-marginal cost relationship.[2]

1. Abba P. Lerner, "The Concept of Monopoly and the Measurement of Monopoly Power," Review of Economic Studies, I (June, 1934), pp. 157-175.

2. The following relationship holds at the short-run equilibrium output:

(i) Harrod. On the basis of this short-run theory, Harrod argued that the fall of real income in the slump would make consumers and business firms more careful in their purchases and make demand more elastic; the trend would be reversed in the boom. Thus, according to him, the degree of monopoly tends to diminish in the slump and increase in the boom.[1]

(ii) Kalecki. Kalecki, however, pointed out importance of other factors, e.g., tacit and explicit collusions among firms, which influence the degree of monopoly in the opposite direction over the cycle.[2] Later, he developed this point in more detail to discuss cyclical and secular behavior of percentage gross margins in American manufacturing.[3] Kalecki began his analysis within the framework of price formation in oligopolistic markets and proceeded to draw implications from actually observed behavior of percentage gross margins. He started with assumptions that manufacturing firms operate with more or less constant marginal and hence average direct cost up to full capacity and that they set their selling prices on the basis of their unit direct costs and their rivals' prices which may be represented by the average price prevailing in the industry. Firms may be said to have higher "monopoly power" if they can charge higher prices with given levels of unit direct cost and industry-average price. Aggregation of micro price-cost relationships over the industry leads to a conclusion that the degree of monopoly power for the industry as a whole can be represented by the price/direct cost ratio, or the average percentage gross profit margin. Kalecki, thus, considered the level of the percentage gross margin as a measure of the degree of monopoly. From this simple formulation of price-cost behavior, Kalecki concluded 1) that increases in the market share of large firms would tend to raise the average degree of monopoly, i.e., the average percentage gross margin, via tacit agreements among them; 2) that sales promotion would tend to create artificial market imperfection so as to raise the degree of monopoly; 3) that rises in the proportion of overhead costs to direct costs would cause rises in the degree of monopoly only insofar as

---

$$m = (1/\eta_d + 1/\eta_c)/(1 + 1/\eta_c)$$

where m is the percentage gross margin, η_d is the elasticity of demand with respect to price, and η_c is the elasticity of supply with respect to average variable cost. When marginal cost is constant, $1/\eta_c$ is equal to zero while m and the degree of monopoly coincide with each other.

1. Roy F. Harrod, "Imperfect Competition and the Trade Cycle," Review of Economic Statistics, XVIII (May 1936), pp. 84-88.

2. Michael Kalecki, Essays in the Theory of Economic Fluctuations, (London, Allen and Unwin, 1939), Chapter 1, "The Distribution of the National Income," reprinted in W. Fellner and B. Haley, eds., Readings in the Theory of Income Distribution, Philadelphia, The Blakiston Co., 1949), pp. 197-217, esp. p. 213.

3. M. Kalecki, Studies in Economic Dynamics, (London, Allen and Unwin, 1943), Chapter I, "Costs and Prices," pp. 9-31, and Theory of Economic Dynamics, (London, Allen and Unwin, 1954), Chapter I, "Costs and Prices," pp. 11-27. Kalecki gave two different versions of essentially the same theory in the two books. We follow above the more recent version of the 1954 book.

"squeezes of profits" lead to tacit agreements among firms--thus, in periods of depression, this "protection of profits" factor would be apt to help create tacit agreements among sellers to recapture cyclical increases in percentage overheads; the degree of monopoly tends to rise in the slump and to fall in the boom; 4) that increased trade union strength may cause firms to adopt lower profit margins to thwart labor's insistent demand for higher wages; 5) that technological progress would influence price-cost relationships only to the extent that increases in overheads due to technological changes are passed on to higher gross margins via tacit agreements among producers.

In the light of these observations, Kalecki interpreted actual movements of the percentage gross margin in American manufacturing as a whole for the period, 1899-1937, on the basis of quinquennial and biennial Censuses of Manufactures. According to him, the secular rise in the percentage gross margin from 1923 to 1929 was partly accounted for by a rapid introduction of sales promotion and a general increase in percentage overheads; rises in the percentage gross margin in the depression, 1929 to 1933, were due to cyclical increases in the percentage overheads which fostered agreements among manufacturers to protect their profits; and declines of the percentage gross margin from 1933 to 1937 were probably due to increased trade union strength.

(iii) Tsiang. Like Kalecki, S. C. Tsiang examined cyclical and secular behavior of the percentage gross profit margin in American manufacturing as a whole and in three selected manufacturing industries for the period, 1919-1937, on the basis of biennial Censuses of Manufactures.[1] Assuming that individual firms have more or less constant short-run marginal cost and that they are faced with kinked demand curves in the short run, he contended that, in explaining cyclical behavior of the percentage gross margin, collusive agreements, tacit and explicit, among firms might be more important than other factors, such as consumers' attitudes (the Harrodian effect) or advertising which would tend to affect demand elasticity. Because of indeterminancy at the kink of the demand curves, Tsiang expected that "it may very well be that the movement of profit margins would not lend itself to generalization into some predictable law, or display any definite pattern of behaviour during the business cycles at all."[2] Examining the biennial time series of the percentage gross margin for manufacturing as a whole, he attached a significance to its rise from 1929 to 1933 in the cyclical downswing and to its decline from 1933 to 1937 during the recovery as an evidence of changes in the intensity of collusions among sellers over the business cycle. For its secular rise from 1921 to 1929, Tsiang disagreed with Kalecki who stressed the role of sales promotion; Tsiang pointed out a negative correlation between changes in the percentage gross margin and changes in average unit prime cost during the period and suggested that price inflexibility due to oligopolistic pricing practices had probably been the major factor for this inverse association between the two. He proposed this interpretation as an evidence of increased price inflexibility in the U. S. economy during the twenties and thirties.[3]

---

1. Sho-Chieh Tsiang, The Variations of Real Wages and Profit Margins in Relation to the Trade Cycle, (London, Pitman and Sons, 1947), Chapters 5 and 6, pp. 68-118.

2. Ibid., p. 75.

3. For a similar approach see John T. Dunlop, "Price Flexibility and the Degree of Monopoly," Quarterly Journal of Economics, LIII (August 1939), pp. 522-533.

Both Kalecki and Tsiang rejected the full-cost principle as an explanation of cyclical and secular behavior of the percentage gross margin on the ground that the full-cost principle, as an ex-post rationalization, does not supply any explanation about how prices are determined.[1]

3. New Empirical Observations. We have examined conflicting views and interpretations of secular and cyclical behavior of the percentage gross margin in U. S. manufacturing before World War II. We shall now turn to our own examination of price-cost behavior in manufacturing as a whole during the period, 1917-1956.

(i) Sources of data. Previous writers, Kalecki and Tsiang, used Censuses of Manufactures as their main source of data. The Census of Manufactures gives data on such items as value of product, cost of materials, and wages and salaries. Direct cost may be defined as the sum of wages and cost of materials. Definitions are clear, but Census data have serious limitations. In the first place, though the Census covers a long period of years, it was not conducted on the annual basis.[2] In the second place, no data on overhead expenses were collected by the Bureau of the Census throughout the period. These limitations allow us little room for a meaningful inspection of cyclical as well as secular behavior of the percentage profit margin.

There is, however, an alternative data source which has not been utilized so far for our kind of studies.[3] It is Statistics of Income, Part II, Corporate Income Tax Returns.[4] It publishes tabulations of income and balance-sheet statements of active corporations by major industry groups for all years since 1916. Statistics of Income data, however, have their own limitations.[5] Nevertheless, a comparison with Census data has demonstrated

---

1. See Kalecki, Studies, p. 27 and Tsiang, op. cit., pp. 71-74. A contrary view was expressed by B. M. Cheek in a similar study of cyclical behavior of the percentage gross margins in Australian manufacturing industries from 1927 to 1946. He emphasized the full-cost principle as an explanation of cyclical stability of the percentage gross margins. See B. M. Cheek, "Economic Theory and Industrial Pricing," The Economic Record, XXV (August 1949), pp. 140-157.

2. It was conducted quinquennially from 1899 to 1919, biennially till 1939, and annually only for the most recent period since 1949 (as Annual Survey of Manufactures). Since 1947 the Bureau of the Census has been following the practice of withholding aggregative data on the value of product and cost of materials from official publication. This adds to the data limitations.

3. See, however, Thor Hultgren, "Cyclical Changes in Costs, Prices and Profit Margins," American Statistical Association, Proceedings of the Business and Economic Statistics Section, 1955-1956, (Washington, D. C., 1957), pp. 191-200. Hultgren based his study of cyclical timing of total cost and net profit margins in manufacturing on the same source as ours.

4. U. S. Treasury, Statistics of Income, Part II, (Washington, D. C., Government Printing Office).

5. Some of these limitations are as follows: First, industrial classification is on a company basis rather than an establishment basis so that activities other than manufacturing are included in manufacturing activities and vice versa. Secondly, fiscal year rather than calendar year is the basis of tax returns. Thirdly, definitions of cost and value of product are different from Census definitions.

that, in the aggregate for manufacturing, the percentage gross margins from the two sources behaved very closely to each other, warranting the reliability of Statistics of Income data for the purpose of the present study.[1]

(ii) Variables defined. First of all, let us define various variables. We mean by gross margins the difference between the value of product and direct or variable cost. Overhead expenses are the difference between total cost (including all costs incurred for production) and direct cost. Net margins or net profits are gross margins minus overheads and before deducting income taxes and dividend payments. These magnitudes, when expressed as a percentage of the value of the product, may be called percentage gross and net margins and percentage overheads. They are represented by m, n and e, respectively, in this study. In addition, capital invested is defined as the value of all assets employed for production; the turnover ratio, i.e., the ratio of capital invested to the value of product, is represented by k in this study and is called the capital/sales ratio or the assets/sales ratio.[2]

(iii) Statistical data. Table 1 presents the percentage gross margin, overheads and other relevant variables for the manufacturing sector as a whole for the period, 1917-1956,[3] some of which are shown graphically on Charts 1 and 2.

To examine the cyclical behavior of the percentage gross margin, we have to first identify various cyclical phases during the period. The relevant variable in this context is a measure of cyclical fluctuations in demand in relation to total capacity in manufacturing, or a measure of excess capacity in proportion to total capacity. Since we do not have any reliable measure of capacity, a device, such as taking a percentage deviation of the actual level of output over the last-peak level of output may be used as a proxy index of excess capacity. The index, thus constructed, is given in column (10).

There is another index which sensitively reflects changes in demand in relation to capacity. It is the proportion of sales by firms making losses to total sales. When demand declines, more firms run into net losses and the proportion of net-deficit sales in total sales rises. This proportion may

---

1. This close correspondence between the two is observable at least since 1933. There are, however, considerable discrepancies before 1931. A detailed examination of the comparability of data is presented in Appendix 1 of my dissertation.

2. These basic definitions, however, have to be modified to make them conform to definitions followed by corporate income tax returns. The value of product is replaced by total sales which is the sum of gross sales of goods (net of discounts and trade allowances) and gross receipts from operations; overheads as all deductions minus direct cost. The percentage gross margin is defined as the ratio of (sales minus direct cost) to sales. The percentage overheads, however, are defined as the ratio of overhead expenses to total receipts (total sales and other income) to make an allowance for non-manufacturing activities which contribute to creating additional income of manufacturing corporations. Correspondingly, capital invested is defined as the book value of total net assets minus holdings of government and other securities. For exact definitions of these concepts, see the glossary provided in Statistics of Income, Part II.

3. No comparable data are available for the first year (1916).

TABLE 1

Percentage Gross Margins, Overheads, Net Margins, Assets/Sales Ratio, Net-Deficit Sales Ratio, and Other Variables, All Manufacturing, 1916-1956

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Annual Changes in						
Year	m	e	n	k	d	Sales	Direct Cost	Overheads	Q	$1-Q/Q_p$	p	w
1916					.068	NA	NA	NA	44	-.19	57.7	$.260
1917	.301	.200	.101		.042	1.79	NA	NA	44	.01	75.9	.310
1918	.254	.185	.069		.051	.05	.12	-.03	43	.02	84.8	.405
1919	.255	.196	.059		.126	.18	.18	.25	38	.14	88.4	.477
1920	.240	.216	.024		.202	.08	.11	.20	39	.11	103.2	.555
1921	.244	.281	-.033		.365	-.32	-.33	-.12	30	.32	66.9	.515
1922	.263	.234	.029		.194	NA	NA	NA	39	.11	63.5	.487
1923	.261	.224	.037		.133	.27	.27	.21	45	-.02	67.0	.522
1924	.235	.215	.020		.160	-.05	-.01	-.08	43	.04	64.4	.547
1925	.220	.202	.018		.131	.11	.13	.06	47	-.04	66.4	.547
1926	.247	.213	.034		.154	.05	.01	.08	50	-.06	65.5	.548
1927	.237	.216	.021		.214	.02	.03	.04	50	.00	62.2	.550
1928	.243	.210	.033		.146	.06	.05	.03	52	-.04	62.7	.562
1929	.247	.207	.040		.171	.08	.07	.06	58	-.12	61.8	.566
1930	.236	.239	-.003		.342	-.19	-.18	-.06	48	.17	57.0	.552
1931	.236	.281	-.044		.521	-.25	-.25	-.12	39	.33	49.6	.515
1932	.238	.355	-.117		.606	-.30	-.30	-.07	30	.48	44.9	.446
1933	.266	.306	-.040		.364	.12	.08	-.05	36	.38	45.6	.442
1933	.276	.283	-.007	1.364	.364	.11	NA	NA	36	.38	45.6	.442
1934	.246	.232	.014	1.087	.329	.16	.21	-.05	39	.33	50.6	.532
1935	.247	.215	.032	.931	.258	.17	.17	.09	46	.21	52.9	.550
1936	.248	.199	.049	.826	.150	.17	.17	.08	55	.05	53.1	.556
1937	.247	.203	.044	.766	.175	.11	.12	.13	60	-.02	56.9	.624
1938	.250	.234	.016	.918	.333	-.18	-.18	-.05	46	.23	53.1	.627
1939	.268	.238	.030	.826	.142	.14	.12	.16	57	.05	52.3	.633

TABLE 1 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Annual Changes in						
Year	m	e	n	k	d	Sales	Direct Cost	Overheads	Q	$1-Q/Q_p$	p	w
1940	.274	.207	.067	.787	.093	.15	.14	.00	66	-.10	53.2	.661
1941	.280	.178	.102	.658	.032	.39	.38	.19	88	-.33	58.1	.729
1942	.264	.158	.108	.606	.032	.29	.32	.14	110	-.25	63.9	.853
1943	.247	.141	.106	.534	.020	.23	.26	.09	133	-.21	64.8	.961
1944	.235	.147	.088	.492	.019	.06	.07	.10	130	.02	65.3	1.019
1945	.232	.171	.061	.507	.055	-.08	-.08	.07	110	.18	66.1	1.024
1946	.251	.181	.070	.592	.075	-.03	-.05	.03	90	.18	75.5	1.088
1947	.248	.166	.082	.533	.048	.31	.31	.20	100	.10	95.9	1.237
1948	.248	.168	.080	.527	.070	.11	.11	.12	103	.06	103.8	1.350
1949	.251	.186	.065	.564	.105	-.07	-.07	.03	97	.14	100.3	1.401
1950	.269	.172	.097	.545	.046	.18	.15	.09	113	-.03	104.1	1.465
1951	.262	.175	.087	.541	.052	.16	.17	.18	121	-.07	115.5	1.59
1952	.265	.189	.067	.566	.064	.02	.03	.10	125	-.03	112.9	1.67
1953	.258	.192	.066	.544	.060	.08	.07	.09	136	-.09	112.8	1.77
1954	.263	.205	.058	.586	.106	-.05	-.05	.02	127	.07	113.7	1.81
1955	.276	.202	.074	.560	.060	.14	.12	.13	140	-.03	115.5	1.88
1956	.277	.211	.066	.591	.061	.04	.04	.09	144	-.03	119.5	1.98

Sources:

Columns (1) to (5): Statistics of Income, Part II, 1916-1956. Col. (1) percentage gross margin; (2) percentage overheads; (3) percentage net margin (= col. (1) - (2)); (4) assets/sales ratio; (5) net-deficit/sales ratio. Due to occasional changes in tabulations, series are not strictly comparable over the entire period. For the period, 1917-1931, variables are computed for all manufacturing corporations reporting income tax returns. For the period, 1932-1956, variables are for balance-sheet reporting corporations only. Differences due to this change are insignificant. For the period, 1922-1931, sales do not include receipts from operations because no separate data were available for the latter in this period. Total receipts, however, include gross profits from operations (gross receipts minus cost of operations. Therefore, the percentage overheads for this period are overestimated. (The series are extended to 1933 to give an overlap with the later period.) For the period, 1917-1921, no separate figures for sales of goods are published, though cost of goods sold is

listed. Sales of goods are estimated by multiplying total receipts with the average ratio of sales to total receipts for the subsequent period, 1922-1929 (except 1925), .9551 (its standard deviation is .0026).

Columns (6) to (8): Annual percentage changes (X_t/X_{t-1}) in total sales (col. (6)), in direct cost (col. (7)), and in overheads (col. (8)). The data are drawn from the same source as above.

Column (9): Federal Reserve Board indexes of production for all manufacturing (1947-9=100). See Productivity, Prices, and Incomes, Materials prepared for the Joint Economic Committee, United States Congress, (Washington, D. C., Government Printing Office, 1957), p. 85.

Column (10): Excess capacity measure. Unity minus the ratio of the actual level of output to the last-peak level of output, Q_p. For 1946 and subsequent years, the 1945 output, instead of the 1943 output, was chosen as Q_p in order to avoid apparent wartime distortion.

Column (11): Wholesale price indexes for total manufactures (1947-9 = 100). For 1947-56, see Wholesale Prices and Price Indexes, 1957, Bureau of Labor Statistics Bulletin No. 1235, (Washington, D. C., Government Printing Office, 1958), pp. 32-33. For earlier years, indexes for semi-manufactured and manufactured goods are combined and linked to the later indexes. These indexes are given in Handbook of Labor Statistics, 1950 edition, BLS Bulletin, No. 1016, (Washington, D. C., Government Printing Office, 1951), p. 118.

Column (12): Average hourly earnings for production workers in all manufacturing. For 1919-1956, see BLS, Employment, Payroll, Hours, and Earnings. For 1916-1918, average weekly earnings given in the same source were divided by average weekly hours worked in manufacturing estimated from Paul Douglas, Real Wages in the United States, 1890-1926, as given in Historical Statistics of the United States, 1790-1945.

---

be called the net-deficit sales ratio or in short the d-ratio. It is given in column (5). Chart 2 demonstrates the close movements of the two measures of excess capacity. [1]

4. Secular Behavior of the Profit Margin. A glance at Chart 1 is enough to demonstrate secular stability of the percentage gross margin for manufacturing as a whole. The percentage overheads, on the other hand,

TABLE 2

Means and Standard Deviations of Percentage Gross Margin and Overheads (%)

Period						
	Means	s.d.	c.o.var.	Means	s.d.	c.o.var.
1922-1933	24.4	1.4	5.7	24.2	4.8	19.6
1933-1956	25.7	1.4	5.3	19.4	3.6	18.1

1. The coefficient of correlation between the two indexes for, e.g., 1922-1943 is 92%.

CHART 1

PERCENTAGE GROSS MARGINS AND OVERHEADS, ALL MANUFACTURING 1917-1956

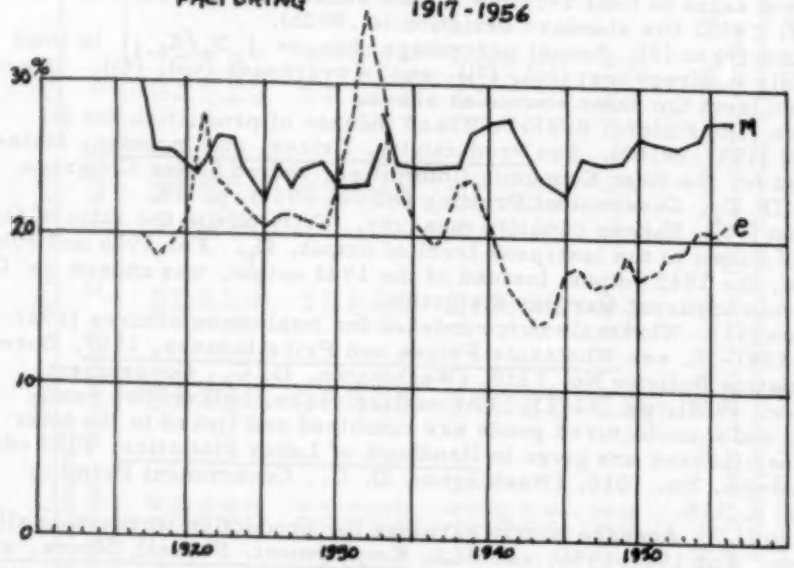
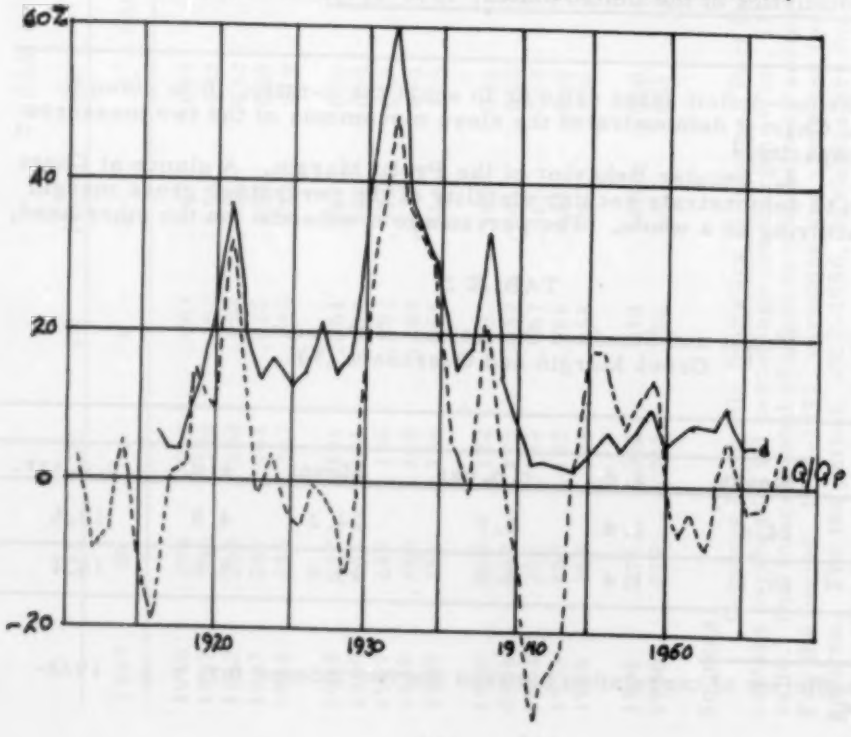


CHART 2

NET DEFICIT SALES RATIO (1916-56) AND EXCESS CAPACITY MEASURE (1911-58) ALL MANUFACTURING



show large fluctuations over the period. Table 2 gives a quantitative expression to this visual impression. The percentage overheads had coefficients of variation about 3.4 times as large as those of the percentage gross margin in both sub-periods.

Overhead expenses, more or less fixed in the short run,[1] fluctuate together with cyclical fluctuations in excess capacity. Cyclical swings of demand are faithfully followed in cyclical movements of the percentage overheads in Chart 1. There, however, exists a remarkable degree of secular stability in the percentage overheads. Their normal level seems to have remained close to 20% through the end of 1930's. Then, from the high of 1939 (23.8%) the percentage overheads declined sharply to the low of 1943 (14.1%) and recovered gradually to the pre-war level by the end of our period. The sharp decline in e from 1939 to 1943 is partly due to the exceptionally high degree of capacity utilization in the wartime. The other responsible factor is a rise in prices of manufactured goods relative to other prices, especially service prices during the same period. The recovery of e since 1943 is partly explainable by the reversal in relative prices after the war and partly by relative increases in overhead labor (salaried personnel).[2]

The percentage gross margin, which was cyclically more stable than the percentage overheads, fluctuated around the normal level of 25 per cent or so. Secular behavior of e since 1939 is closely but less markedly reflected in m , too. Wartime excess demand and sharp increases in corporate income tax rates from 1940 apparently contributed towards keeping the percentage gross margin more stable in the face of declining overheads.

5. Cyclical Behavior. There, however, seems to be a consistent, even though not conclusive, pattern in the cyclical behavior of m . We may hazard a generalization from Charts 1 and 2; the percentage gross margin rises somewhat in a year of recovery and then tends to decline gradually to the more or less stable floor (relative to the normal level of e) through the ensuing prosperity and contraction. Cyclical behavior of m and e in years of recovery and contraction, along with that of wholesale price indexes and average hourly earnings, is tabulated in Table 3 to make this general pattern clearer.[3]

Notable exceptions from our general rule arise only in 1923-25. Data for the early 20's are less reliable than later years and these exceptions should probably be neglected. It should be noted that constant m's in recent years of decline were actually falls as compared to the rising e's in the same period. It must also be noted that the cyclical behavior of m is not necessarily

---

1. For annual percentage changes in total sales, direct cost, and overhead expenses, see columns (6), (7), and (8) on Table 1. For the 37-year period, 1919-1956, their averages are almost equal to each other, being 6.4, 6.4, and 6.6% respectively. However, their standard deviations are quite different, being 16.1%, 16.4%, and 9.3%. The coefficients of correlation are .989 between changes in sales and in direct cost, and only .750 between changes in sales and in overheads.

2. For this substitution of overhead labor for direct labor, see Ruth P. Mack, "Inflation and Quasi-Elective Changes in Costs," Review of Economics and Statistics, XLI (August 1959), pp. 225-231

3. It should be noted that this particular pattern is not due to cyclical shifts in industry composition within the manufacturing sector.

TABLE 1

Cyclical Behavior of the Percentage Gross Margin, Overheads, Prices and Average Hourly Earnings for All Manufacturing, 1917-1956 [1]

Recovery					Decline				
Year	m	e	p	w	Year	m	e	p	w
1921-22	rise	fall	fall	fall	1918-19	const.	rise	rise	rise
1924-25	fall	fall	rise	const.	1923-24	fall	fall	fall	rise
1927-28	rise	fall	const.	rise	1926-27	fall	const.	fall	const.
1932-33	rise	fall	rise	const.	1920-30	fall	rise	fall	fall
1938-39	rise	const.	fall	rise	1937-38	const.	rise	fall	const.
1946-47	const.	fall	rise	rise	1944-45	const.	rise	rise	const.
1949-50	rise	fall	rise	rise	1948-49	const.	rise	fall	rise
1954-55	rise	const.	rise	rise	1953-54	const.	rise	rise	rise
Number of years									
rise	6	0	5	5		0	6	3	4
fall	1	6	2	1		3	1	5	1
const.	1	2	1	2		5	1	0	3

due to particular patterns of cyclical behavior of p and w.

For years other than those of turning points, the percentage gross margin remained more or less stable, or gradually falling. Such are years, 1930-32, 1934-38, 1941-45, 1946-48, 1950-54, and 1955-56. Deviations are found for the following years:

Rise: 1925-26, 1945-46
Fall: 1917-18, 1919-20, 1933-34.

The fall of m from 1917 to 1918 can be explained as a recovery from the post-World War I inflation to more normal price-cost relationships. Though comparable data are not available for 1916, there is reason to believe that m rose sharply from 1916 to 1917 due to very strong excess demand. From 1916 to 1917 the wholesale price index for manufactured goods rose by 31.5% whereas average hourly earnings rose only by 19.2%.[2] A similar demand inflation of the classical type, though milder, took place in 1945-46 when the wholesale price index rose by 14.2% and average hourly earnings by 6.1%. The exceptions in 1920's might be due to defects in data. The fall of m in 1933-34 is a reaction to the sharp rise of m in the preceding recovery of 1932-33.

6. An Interpretation. We may summarize our findings as

1. For m and e, changes within $\pm$ 2.5 per cent and for p and w, changes within $\pm$ 1 per cent are treated as constant.

2. From 1917 to 1918, p rose by 11.7 per cent whereas w increased by 30.7 per cent.

follows:[1] Let s and c be subscripts representing the secular and cyclical components of the percentage ratios.[2] m_c is positive in a year of recovery and then decreases gradually over the cycle,[3] while e_c is at minimum in the cycle peak and increases as the degree of excess capacity increases. The variability of m_c was about one third of that of e_c in the past four decades and there exists a strong negative correlation between m_c and e_c.[4]

Now how did the secular components of m and e change over the four decades? The fact that the large swing of e_s since 1939 is closely reflected in a similar movement of m_s suggests that changes in the secular level of e are most likely to be passed on to the secular level of m with very little time lag. It is also apparent that the secular component of the percentage net margin, m_s, expanded greatly from 1940 on. Part of this widening is due to a substantial decline in the degree of excess capacity during the wartime and postwar period. But it also corresponds in timing with the large jump in the corporate income tax rates since 1940.[5] The question

---

1. Annual data on which our findings are based are not a satisfactory basis for examining cyclical behavior. Comparable quarterly data are available in FTC-SEC, Quarterly Financial Report of Manufacturing Corporations from the first quarter of 1947 on. Unfortunately, however, it does not segregate direct cost and overheads except for the nine quarters beginning the first quarter of 1951, thus making a study of quarterly movements of m impossible.

2. By definition,

$$m = m_s + m_c \quad \text{and} \quad e = e_s + e_c$$

3. According to B. M. Cheek, Australian manufacturing in the period of 1927-46 showed an exactly opposite pattern of cyclical behavior of m -- rising in recession and falling in recovery. This is rather a puzzling difference. See Cheek, op. cit.

4. If linear trends are fitted to m and e during the period, 1947-56, to estimate their cyclical components, we obtain

$$m_s = .29t + 24.8 \quad (R^2 = .67)$$
$$e_s = .49t + 16.4 \quad (R^2 = .86)$$
$$r(m_c, e_c) = -.52$$

During the 1947-56 period the normal net margin continued to narrow.

5. The effective tax rate on net profit in all manufacturing for 1938-1952 was as follows (expressed in per cent):

1938	15.5	1943	63.0	1948	36.0
1939	16.0	1944	62.5	1949	35.8
1940	27.6	1945	57.9	1950	43.9
1941	46.7	1946	36.5	1951	55.6
1942	59.6	1947	36.1	1952	54.6

For 1954-56 the rate was reduced to about 47 per cent, as the excess profits taxes during the Korean war were abolished. See Eugene M. Lerner and Eldon S. Hendriksen, "Federal Taxes on Corporate Income and the Rate of Return on Investment in Manufacturing, 1927 to 1952," National Tax Journal, IX (September 1956), pp. 193-202.

of whether or not corporate income taxes are shifted forward has been discussed by a number of writers. Traditional theory argues that, as corporate income taxes change neither marginal revenue nor marginal cost curves of the firm, they will not affect the profit maximizing price-output relationships and hence cannot be shifted. This argument is valid only in the short run when marginal revenue and marginal cost curves are given. In the longer run, changes in tax rates affect the relative size of earnings of individual firms and thus tend to reallocate resources among firms and to affect relative growth and new entry of firms. The distribution of marginal cost and marginal revenue curves among firms will be altered. The statistical evidence of forward shifting of corporate income taxes can only be explained in terms of the particular dynamic adjustment of firms to the changing economic environment.[1]

The same argument may be applied to the effect of secular changes in the proportion of overhead costs to direct costs in so far as these changes come from sources external to individual firms. Changes in general technology or relative prices are external forces which alter the basic cost conditions of individual firms alike. If aggregate demand is more or less normal, changes in their percentage overheads tend to reallocate flows of funds among firms and thus to bring about certain readjustments of aggregate price-cost relationships. Cyclical fluctuations in their percentage overheads, however, do not affect the percentage gross margin in a similar way because all firms are held more or less unchanged.

7. Critique of Previous Studies. Our findings are in sharp contrast to the previous studies by Kalecki and Tsiang in many respects. We have found that, according to our data, there was no secular rise in the percentage gross margin for manufacturing as a whole during the twenties. Even if this absence of the secular trend were due to possible defects in data (e.g., in classifying costs between direct and overhead costs), it still holds that the secular components of the percentage gross margin and the percentage overheads moved closely together. Thus, both Kalecki's argument for selling promotion through intensified advertising campaigns as partly responsible for the increase in the percentage gross margin in the 20's and Tsiang's allegation of increased price inflexibility through the 20's cannot be accepted. We have also found that the percentage gross margin, when cyclical fluctuations were eliminated, remained stable till late in the 30's. Kalecki's inference that increased trade union strength was responsible for the decline in the percentage gross margin over 1933-1937 is therefore not acceptable. We have found that the percentage gross margin showed a more or less consistent pattern of cyclical behavior over the four decades; in particular, little supporting evidence was found for the tendency of the percentage gross margin to rise in periods of depression because of tacit and explicit collusions among sellers. Nor can changes in demand elasticity alone account for it. Finally, we have found that the secular decline of the percentage overheads from 1939 to 1943 and their secular advance since 1943 (and the sharp increase in income tax rates since 1940) were followed closely by the simultaneous parallel movement of the percentage gross margin. This fact alone seems to disprove Kalecki's assertion that changes in normal overheads are passed to gross margins only indirectly through tacit agreements among firms. He did not take into account dynamic factors which may tend to establish a certain stable price-cost structure.

---

1. See in this respect arguments presented by Lerner and Hendriksen, op. cit., and Don M. Soule, "Shifting of the Corporate Income Tax: A Dynamic Analysis," Journal of Finance, XIV (September 1959), pp. 390-402.

Harrod emphasized changes in demand elasticity as a major determinant of the price-cost relationships in the short run. It, however, was recognized that firms tend to attempt long-run profit maximization rather than short-run profit maximization. Also it was doubted that firms can estimate their short-run elasticity of demand in the fact of constantly changing demand conditions. If demand elasticity has something to bear on the percentage gross margin, it should be its longer-run normal aspect.

Kalecki's model, though ingenious, does not provide us with a proper test of his hypothesis. A rise in the percentage gross margin may be due to an increase in the degree of market imperfection but it may be due to any one of other equally important factors.[1] Unless we can test it empirically, we are always led to tautologies.

Tsiang's contention on unpredictableness in cyclical and secular behavior of the percentage gross margin is not borne out by evidence. Reliance on kinked demand curves is not a proper way of explaining the working of our price system over the long run.

The full-cost principle likewise does not tell us why producers set their prices at their actual level and gives us no principle to explain cyclical and secular behavior of the percentage gross margin. It is only a rule of thumb that manufacturers follow. It is perfectly compatible with the marginal principle if we take a longer-run view.

8. Conclusion. We should be able to fully explain the observed cyclical and secular behavior of the percentage gross margin for manufacturing as a whole on the basis of the traditional marginal principle if individual firms have more or less constant marginal cost curves in the short run and if the distribution of demand among firms relative to each other does not change significantly over time. Suppose that individual firms were to set their prices by adding certain percentage markups on average unit direct cost as a short-cut approximation to profit maximization. These markups would be changed only when there were substantial changes in demand and/or cost conditions. Then, in the short run, the average percentage gross margin for manufacturing as a whole would remain more or less stable unless aggregate demand went beyond the range of constant marginal costs or the distribution of demand among firms as well as among industries changed in a significant manner. The average percentage markup or gross margin would tend to fluctuate around the stable level as demand expanded and contracted cyclically, following the common dictum of "what the market will bear." As the boom progressed, productive capacity would tend to outstrip aggregate demand, which would tend to narrow the percentage markup by way of intensified competitiveness among firms. In recession or depression, individual firms would try to maintain their conventional markups the best they could. However, once revival came, their expectations of improved business would tend to raise their markups. Now the markups themselves would be determined by the dynamic interactions of demand and supply. Changes in income tax rates or technological changes which alter the proportion of overheads to direct costs would have the strongest effects on firms marginal to the industry. There would ensue a new distribution of demand among firms and a new structure of percentage markups among firms. The average percentage gross margin for manufacturing as a whole would tend to be realigned to new cost conditions. This brief sketch seems to explain the actual patterns of cyclical and secular behavior of the percentage gross

---

1. Similarly, Lerner's degree of monopoly power can be a measure of monopoly power in the short run only if firms maximize their short-run profits with no reference to their long-run position.

margin for manufacturing as a whole.

B. Intersector Comparison

1. Problem. In the previous section we have argued that the cyclical and secular behavior of the percentage gross margin in relation to the percentage overheads in the manufacturing sector as a whole can be explained for the most part by the marginal principle provided that individual firms have more or less constant marginal cost and that the distribution of demand among firms is stable.[1] Secular changes in the percentage gross margin are ascribed to long-run adaptations of supply to demand under the influence of the marginal principle.

The same hypothesis may be tested in a broader framework by comparing market performances of various sectors with respect to cyclical and secular behavior of the percentage gross margin in relation to the percentage overheads.[2] We shall make such a comparison in this section. The sectors which are selected for the comparison are: farm, mining (including quarrying), wholesale and retail trade, transportation, construction, and services. The data for these sectors are taken from the same source as for manufacturing (except for the farm sector). The definitions of the variables are also the same. The percentage gross margin, percentage overheads, and net-deficit sales ratio are shown graphically in Chart 3.[3] To assist the comparison, means and standard deviations of m and e are shown in Table 4A and those of annual changes in sales, direct cost, and overheads in Table 4B.

2. Empirical Examination. Only a very brief examination of each sector's performance is presented below. The purpose of this section is not to present a rigorous analysis of price behavior but to examine to what extent the cyclical and secular patterns of the profit margin observed for manufacturing are typical for other sectors in the economy. The examination is by necessity sketchy and must be supported by further investigations.

(i) Farm.[4] The farm sector has an inelastic supply in

---

1. As shown in the previous section, the percentage gross margin is a function of the elasticity of demand and the average cost-marginal cost relationship if short-run profits are maximized by individual firms. If the elasticity of demand is taken to be more or less stable in the short run, then the percentage gross margin is dependent mostly on the elasticity of the average-cost function. If the distribution of demand among firms is also stable in the short run and if most firms in the particular sector have similar AC functions, the same holds true with the sector-average percentage gross margin. A very high correlation between annual changes in sales and direct cost (99.5% as shown on Table 4B) suggests the working of manufacturing firms within the horizontal range of average cost curves.

2. The approach is suggested by Richard Ruggles, "The Nature of Price Flexibility and the Determinants of Relative Price Changes in the Economy," in Business Concentration and Price Policy, A Conference of the Universities-National Bureau Committee for Economic Research, (Princeton, N. J., Princeton University Press, 1955), pp. 441-495.

3. For the underlying data of this section see Statistical Appendix of my dissertation.

4. Data for the farm sector are taken from Survey of Current Business, XXXIV, (August 1954), pp. 18-24, and U. S. Income and Output, A Supplement to Survey of Current Business, 1959, pp. 226-227. See also Agricultural Statistics, 1957 edition, Table 666, and 1958 edition, Table 187.

TABLE 4

A. Means and Standard Deviations of Percentage Gross Margins and Overheads, 1933-1956 and 1938-1956

	Gross Margin			Overheads			r(m, e)
	Means	s.d.	c.o.var	Means	s.d.	c.o.var	
A. 1933-1956							
Manufacturing	25.6	1.4	5.3	19.4	3.6	18.6	.31
Farm	41.0	4.4	10.7	18.2	4.0	21.8	.29
Mining	35.6	3.3	9.4	30.4	9.4	31.1	.13
Trade	20.5	.9	4.5	18.5	1.7	9.3	.35
Construction	18.4	2.5	13.7	17.0	5.3	32.2	.95
Services	49.1	6.8	13.8	47.1	11.1	23.5	.93
B. 1938-1956							
Manufacturing	25.6	1.4	5.3	18.5	2.6	14.2	.46
Farm	40.1	3.7	9.2	17.0	3.1	18.0	-.07
Mining	35.6	3.4	9.5	28.8	3.8	13.2	.58
Wholesale Trade	14.1	.4	2.6	12.0	.6	4.9	-.15
Retail Trade	27.2	1.9	7.0	24.3	1.4	5.6	.41
Transportation	32.3	4.4	13.7	26.4	4.0	15.2	.18
Construction	17.4	1.4	8.1	14.9	2.0	13.5	.65
Services	46.1	3.5	7.6	42.0	4.6	10.8	.79

the short run and cyclical changes in demand are absorbed primarily by changes in prices. Thus, we should expect an inverse year-to-year movement of the percentage overheads and a positive movement of the percentage gross margin in relation to that of prices. This expectation can be confirmed by Chart 3A. It must be observed that the percentage gross margin remained within fairly stable ranges in spite of very large price fluctuations. Part of this stability is due to the fact that part of agricultural output is fed back into production as inputs.[1] For the long-run adjustments of supply to demand we should expect that the marginal principle would remain imperfect in its working even over a long period of years when excess supply persists because of a fairly

---

For the farm sector, sales are defined as the sum of receipts from marketings, CCC loans, and home consumption; gross margins as sales minus intermediate products consumed and employee compensation. The ratio between the two is marked by m in Chart 3A. To take account of family labor, imputed compensations for family labor are deducted from gross margins. The ratio of the adjusted gross margins to sales is marked by m' in the chart. (The imputation was made on the assumption that family labor, including proprietors', received the same average compensation per head of hired labor.) Overhead expenses are the sum of gross rents paid to non-farm landlords, other items, capital consumption allowances, and indirect business taxes. Prices and output indexes refer to total sales as defined above.

1. The coefficient of correlation between annual changes in sales and direct cost is 78.5% as shown in Table 4B.

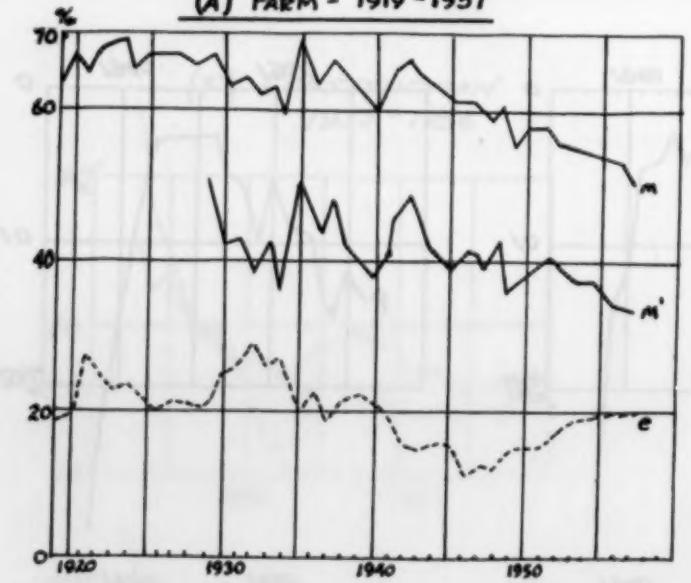
TABLE 4

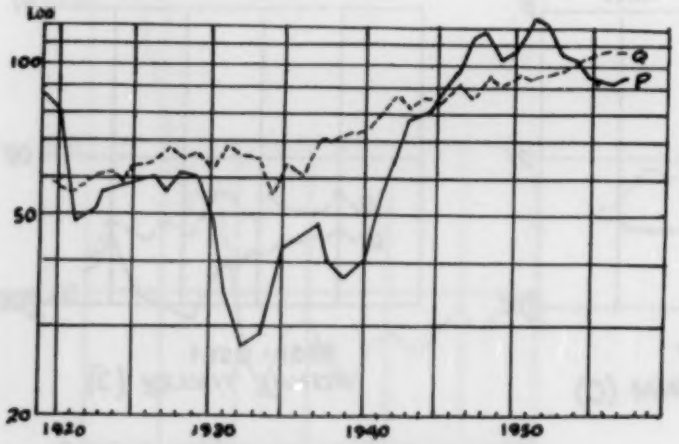
B. Annual Changes in Sales, Direct Cost, and Overheads

Sector	Period	Average annual changes (%) in*			Correlation between changes in	
		Sales	direct cost	overheads	Sales & direct cost	Sales & overheads
Manufacturing	1918-1956	6.4 (16.1)	6.4 (16.3)	6.6 (9.4)	.995	.762
Farm	1920-1956	3.2 (16.9)	4.1 (15.2)	2.4 (8.9)	.785	.381
Mining	1933-1956	8.8 (14.4)	7.8 (13.1)	7.2 (11.9)	.969	.827
Wholesale Trade	1938-1956	11.1 (13.2)	11.3 (12.9)	10.7 (10.5)	.998	.898
Retail Trade	1938-1956	10.9 (10.8)	11.2 (11.7)	9.7 (8.5)	.991	.941
Transportation	1938-1956	8.1 (9.0)	8.1 (7.0)	6.3 (7.5)	.805	.423
Construction	1932-1956	16.6 (23.1)	17.3 (21.0)	11.1 (14.3)	.892	.743
Services	1932-1956	9.0 (9.9)	10.5 (8.3)	6.3 (12.1)	.782	.830

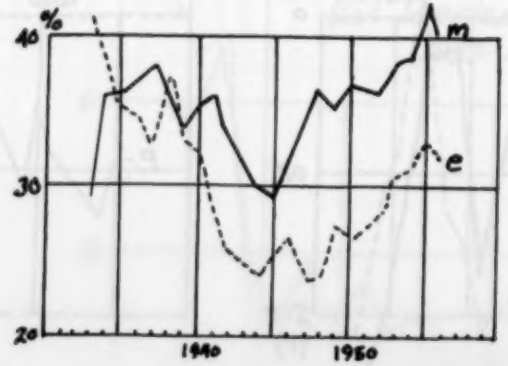
* Numbers in parentheses represent standard deviations.

(A) FARM - 1919 - 1957





(B) MINING 1933 - 1956



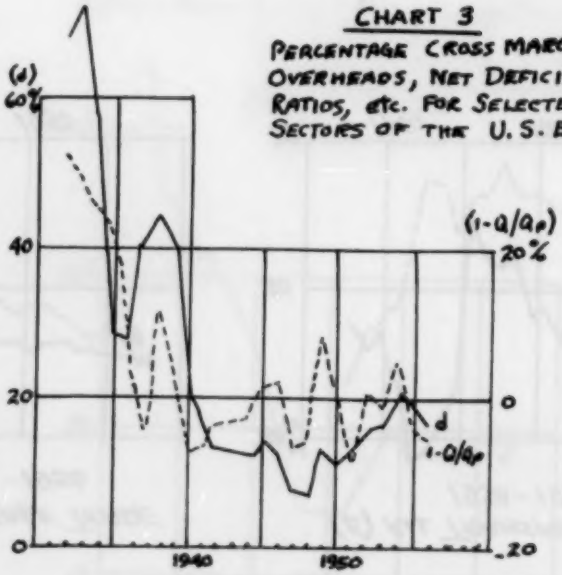
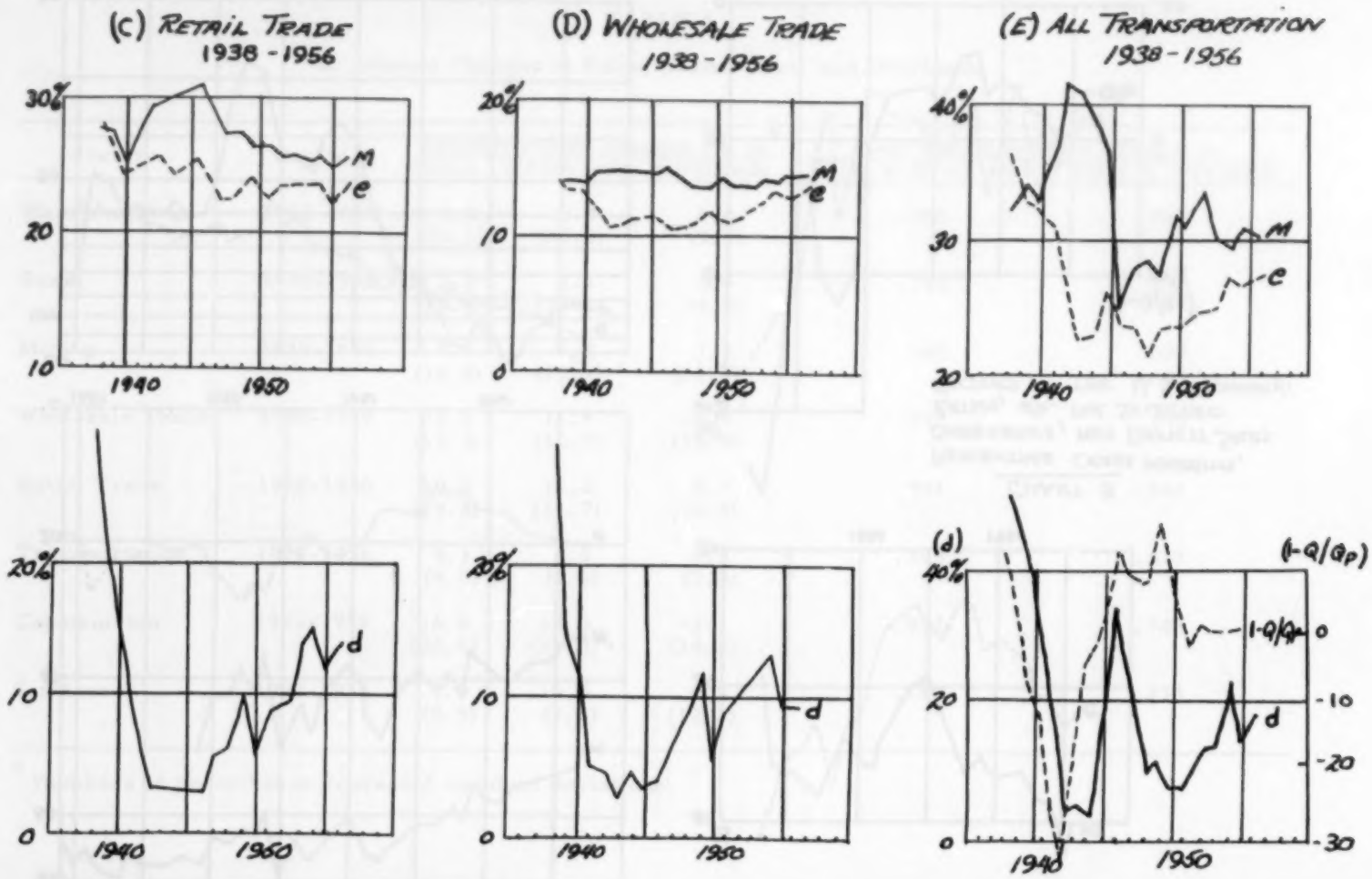
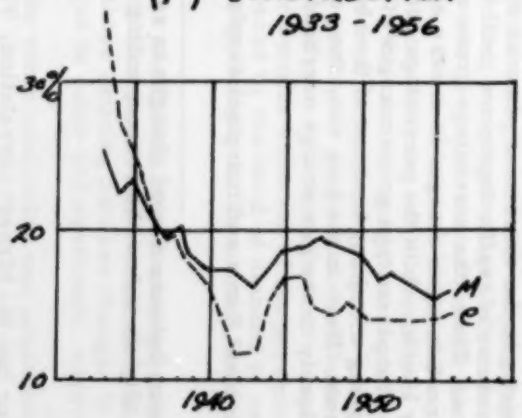
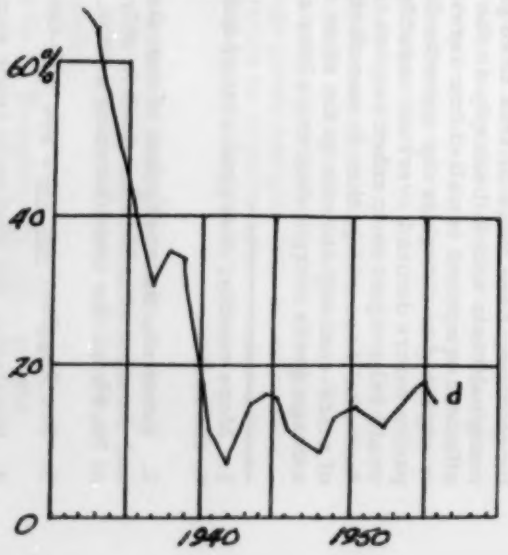


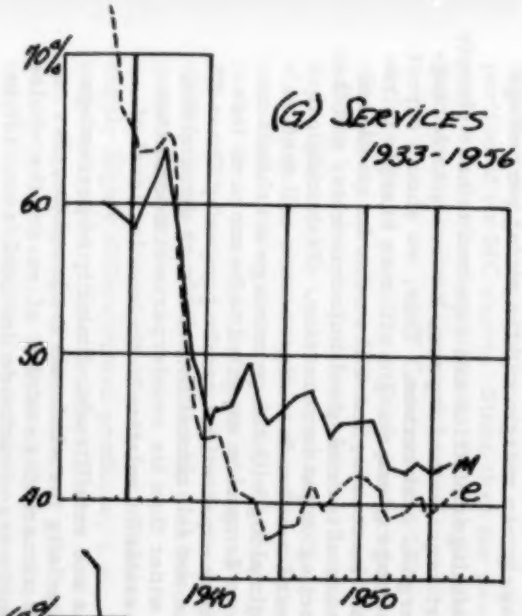
CHART 3

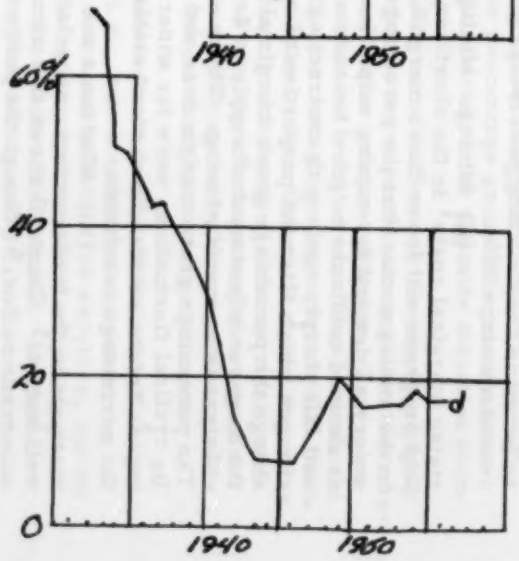
PERCENTAGE CROSS MARGINS, OVERHEADS, NET DEFICIT SALES RATIOS, etc. FOR SELECTED SECTORS OF THE U. S. ECONOMY.


(c) Retail Trade
1938 - 1956
30%
20
10
1940
1950
M
e
(D) Wholesale Trade
1938 - 1956
20%
10
0
1940
1950
M
e
(E) All Transportation
1938 - 1956
40%
30
20
1940
1950
M
e
20%
10
0
1940
1950
d
20%
10
0
1940
1950
d
(d)
40%
20
0
1940
1950
1-Q/Qp
d
0
-10
-20
-30
(1-Q/Qp)


(F) CONSTRUCTION
1933 - 1956
30%
20
10
1940
1950
M
e


60%
40
20
0
1940
1950
d


(G) SERVICES
1933-1956
70%
60
50
40
1940
1950
M
e


60%
40
20
0
1940
1950
d

wide gap between the gross margin and overheads. Farm output grew steadily through the 1940's and the 1950's, especially since 1951. As demand did not follow supply, farm prices tended to decline continually since 1951 in spite of Governmental price supports; the imperfect adaptation of supply to demand is reflected in the declining percentage gross margin and the rising percentage overheads since 1951.

(ii) Mining. Mining industries are expected to have sharply rising marginal costs[1] in the short run. In the long run, however, technological progress will lower these marginal cost curves. Thus, we should expect in the mining sector that the percentage gross margin will rise sharply in recovery as demand for mining output increases, but will tend to be stabilized as demand continues to grow because of technological improvements, and then will fall sharply again with contracting output in recession. Technological changes which alter the proportion of overheads to direct costs will tend to change the percentage gross margin along with the percentage overheads as the long-run adjustment of supply to demand is expected to be more or less efficient as in manufacturing. These expectations can be proven by Chart 3B. The percentage gross margin rose and fell almost invariably as demand did. Its cyclical fluctuations were far wider than its counterparts in manufacturing.[2] Yet its secular level stood stable in relation to the secular level of the percentage overheads.

(iii) Wholesale and retail trade. Stability of percentage markups in the trade sectors, especially the wholesale trade sector, is well known.[3] Chart 3D shows this remarkable stability of m for the wholesale trade sector.[4] Though the percentage overheads declined from 1938 to 1942 and rose very gradually since then, the very strong wartime demand (and rises in income tax rates) and then the postwar return to a higher degree of competitiveness acted as an offset. In the retail trade sector no secular trend in the percentage overheads is observable, probably because technological changes are the least strong in this sector for replacing labor. Secular strength of demand is solely reflected in the behavior of the percentage gross margin over the period. In both sectors, the percentage gross margin seems to be insensitive to cyclical fluctuations in demand.

(iv) Transportation. Transportation industries, as public utilities, have more or less fixed prices in the short run, whereas their marginal costs will fall sharply as the volume of traffic increases until less efficient equipment is called into service. Thus, the percentage gross margin is expected to change in the opposite direction from changes in excess capacity before demand overruns capacity limits, while the percentage overheads would behave just as in other sectors. The percentage gross margin will fluctuate more widely than in manufacturing or mining because of the absence of price-cost adjustments in the short run. Even in the long run, the percentage gross margin may move less closely to the percentage overheads than

---

1. More exactly, marginal cost of individual firms and marginal supply price to the sector.

2. However, the coefficient of correlation between annual changes in sales and direct cost is 96.9%, which is only slightly below the corresponding figure of 99.5% for the manufacturing sector.

3. See Ruggles, op. cit., pp. 485-486.

4. Similarly, the wholesale trade sector has the highest correlation, 99.8%, between annual changes in sales and direct cost.

in manufacturing because of time lags in price-cost adjustments. Chart 3E supports these expectations. We should note a sharp rise in m from 1940 to 1943 as demand recovered and then a sharp fall in m from 1943 to 1946 as demand contracted. This movement of m is compatible with the assumption that the transportation sector was at work along the downward sloping portion of its MC curve. Since then the percentage gross margin was more stable as demand stabilized. But it still moved in the opposite direction from fluctuations in the degree of excess capacity.

(v) Construction. The construction sector supplies highly differentiated products. Each product has to meet particular specifications demanded by customers. Detailed cost estimates are prerequisite. Cost-plus pricing based on both direct and overhead costs is followed. Thus, we should expect that in the construction sector the percentage gross margin would move closely to the percentage overheads over the cycle. Chart 3F shows the close association between m and e . This cyclical behavior is diametrically different from manufacturing or mining. From the chart we can also see that since 1940 the gap between m and e was influenced by the conditions prevailing in the market. It must be pointed out that the construction sector is quite competitive.

(vi). Services. This sector consists of industries which supply highly differentiated products, i.e., services of various sorts, to both business and households. Prices are hard to define for services, but we can reasonably expect that full-cost pricing may be the prevalent pricing policy in the service sector as it is in the construction sector. Markets may be more differentiated than in any other sector and price discrimination more common. Besides, costs which are classified as overheads may be considered as part of direct cost in this sector. Chart 3G presents remarkably close cyclical movements of the percentage gross margin and overheads in this sector.[1] On the other hand, the secular level of the percentage net margin is again apparently governed by strength of demand.

3. Conclusion. Through this brief examination of the seven sectors, it is apparent that differences in cyclical behavior of the percentage gross margin among the sectors can be explained for the most part by differences in shapes of marginal cost curves in these sectors. The percentage gross margin exhibits varying degrees of sensitivity to cyclical changes in demand, ranging from positive reactions in the farm sector and the mining sector to negative reactions in the construction sector and the services sector. No reference to industrial concentration is required to explain the differences in these cyclical patterns of behavior.[2]

We can then safely conclude that the particular pattern of cyclical behavior of the percentage gross margin in manufacturing is most likely due to the particular shape and distribution of its marginal cost curves. The marginal principle is sufficient to explain this particular cyclical behavior. As for secular behavior of the percentage gross margin, it is more or less influenced by the level of demand in relation to capacity after secular adjustments to changes in the normal level of the percentage overheads.

---

1. The coefficient of correlation between annual changes in sales and the sum of direct cost and overheads is 93.7%, which is substantially higher than those existing between changes in sales and direct cost, and between changes in sales and overheads, separately.

2. The same conclusion was presented by R. Ruggles, op. cit., p. 486.

C. Size of the Firm and Price-Cost Structure

1. Problem. We have found that the percentage gross margin in manufacturing as a whole exhibited a particular pattern of cyclical and secular behavior. It may be asked whether or not this result is independent of aggregation. Do firms of different sizes behave in a similar manner? Are there any differences in the behavior of individual industries in this respect? We examine the first question in what follows.

2. Intersize Differences in Price-Cost Structure. Statistics of Income divides firms into a number of classes by the size of their total net assets.[1] Based on these size classes, changes in percentage gross margins and overheads were compared for the period, 1947-1956. Though there are some variations from year to year, m and e in all size classes tended to change in the same direction as the aggregative averages did.

Chart 4 illustrates this relation for the most recent period, 1954-56. From 1954 to 1955 the recovery of m was faster in larger size classes, but from 1955 to 1956 the slight increase of m for manufacturing as a whole was accompanied by more or less even increases of m in all size classes (except for the largest size class). It suffices here to present a general conclusion that large firms do not seem to be less sensitive to cyclical changes in demand. Chart 4, in addition, reveals that there is a very clear relationship between size of the firm and price-cost structure and that the net-deficit sales ratio declines steadily as size of the firm increases. To the examination of these facts we now turn.

Statistics of Income divides corporate firms into those which made profits and those which made losses, in addition to size classes. Chart 5 shows graphically the percentage gross margin, percentage overheads, capital/sales ratio, and rates of return (before tax) on capital invested for the net-income and net-deficit groups over all size classes in manufacturing for the year 1954.[2]

Both the net-income and net-deficit groups show parallel movements of m and e, falling sharply and then rising again, and the gap between them tends to increase (algebraically) as size of the firm increases. The capital/sales ratio tends to rise steadily as corporate size increases. Rates of return for the net-income group tended to be fairly stable over size classes.[3] Capital invested in manufacturing had only slightly increasing rates of return in 1954 as size of the firm increased. The net-deficit group, however, showed a large variability of rates of return in spite of more or less stable percentage net losses because of rising capital/sales ratios with corporate size.

---

1. For sizes of these classes, see Table 5 below.

2. For the underlying data see Statistical Appendix of my dissertation.

3. Rates of return before tax on capital invested ranged from 9.4% (Class IV) to 12.6% (Classes IX and X). The maximum range after tax is one half of this difference, i.e., 1.6%.

For previous studies of corporate size and earning power, see William L. Crum, Corporate Earning Power, (Palo Alto, California, Stanford University Press, 1929), and Corporate Size and Earning Power, (Cambridge, Mass., Harvard University Press, 1939); Joseph L. McConnell, "Corporate Earnings by Size of Firm," Survey of Current Business, XXV (May, 1945), pp. 6-12, and "1942 Profits by Size of Firm," Survey of Current Business, XXVI (January 1946), pp. 10-16; Sydney Alexander, "The Effect of Size of

CHART 4. ALL MANUFACTURING BY ASSET SIZE CLASS 1954 - 1956

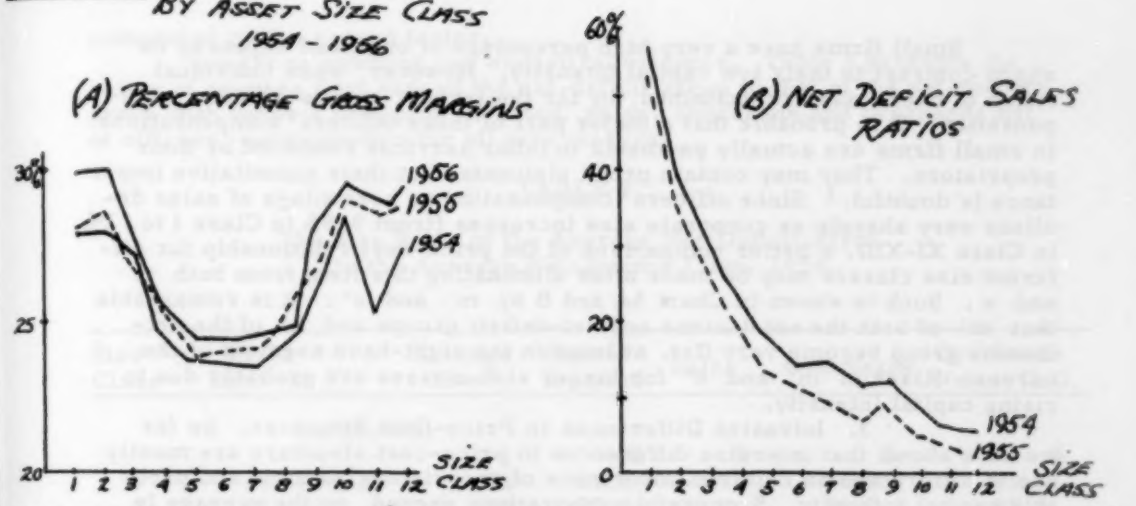
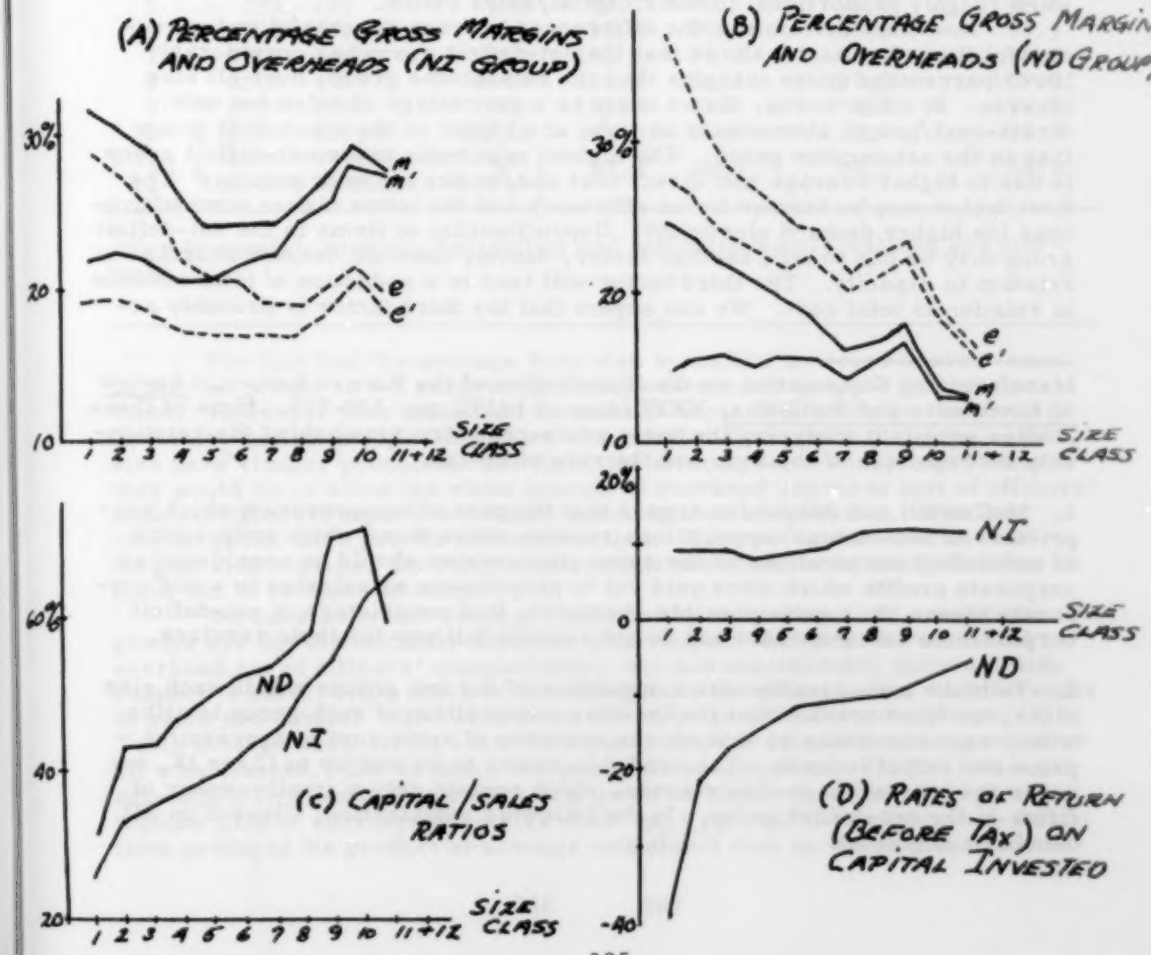


CHART 5. NET INCOME (NI) & NET DEFICIT (ND) MANUFACTURING CORPORATIONS BY ASSET SIZE CLASS, 1954



Small firms have a very high percentage of overhead expenses in sharp contrast to their low capital intensity. However, when individual items of overheads are examined, by far the largest item is officers' compensation. It is probable that a major part of these officers' compensations in small firms are actually payments to labor services rendered by their proprietors. They may contain profit elements, but their quantitative importance is doubtful.[1] Since officers' compensation as percentage of sales declines very sharply as corporate size increases (from 9.5% in Class I to .2% in Class XI-XII), a better comparison of the price-cost relationship for different size classes may be made after eliminating this item from both m and e. Such is shown in Chart 5A and B by m' and e'. It is remarkable that m' of both the net-income and net-deficit groups and e' of the net-income group become very flat, at least in the right-hand segment of the curves. Rises in m' and e' for larger size classes are probably due to rising capital intensity.

3. Intrasize Differences in Price-Cost Structure. So far we have shown that intersize differences in price-cost structure are mostly due to differences in relative importance of proprietors' income and probably capital intensity. Successful corporations earned, on the average in the postwar period, only slightly increasing rates of return on capital invested as corporate size grew. In other words, their percentage net margins were roughly proportional to their capital/sales ratios.

Now what determines the differences between successful and unsuccessful firms? Chart 5 shows that the net-deficit group has considerably lower percentage gross margins than the net-income group, over all size classes. In other words, direct costs as a percentage of sales (or unit-direct-cost/price, abbreviated as c/p) are higher in the net-deficit group than in the net-income group. The higher c/p ratio in the net-deficit group is due to higher average unit direct cost and/or due to lower prices.[2] The first factor may be termed lower efficiency and the latter higher competitiveness (or higher demand elasticity). Unprofitability of firms in the net-deficit group may be due to still another factor, namely absolute demand shortage in relation to capacity. The third factor will lead to a reduction of total revenue in relation to total cost. We can expect that the third factor is probably ac-

---

Manufacturing Corporation on the Distribution of the Rate of Return," Review of Economics and Statistics, XXXI (August 1949), pp. 229-235. None of these studies goes into analyzing the basic price-cost structure behind the relationship between size of the firm and the rate of return.

1. McConnell and Alexander argued that the part of compensation which proprietors of net-income corporations receive above those which proprietors of net-deficit corporations of the same size receive should be considered as corporate profits which were paid out to proprietors as salaries to avoid corporate taxes. It is quite possible, however, that proprietors of net-deficit corporations are underpaid and do not receive full pay for their services.

2. To make a reasonably safe comparison of the two groups within each size class, we must assume that the industry composition of each group is alike within each size class so that we can conceive of some sort of aggregative price and output indexes. The condition seems to be met up to Class IX, but not in the largest three size classes, which contain only a small number of firms in the net-deficit group. In the following examination, Class X to XII should be neglected.

companied by the second factor.

Now let us examine our statistical evidence. Table 5 presents the ratios of average firm size of the net-deficit group to that of the net-income group with respect to capital invested, sales, overheads and overheads net of officers' compensation for all size classes.

TABLE 5

The Size Ratio of the Average Net-Deficit Firm to the Average Net-Income Firm, 1954

Size Class	Class Interval	Average Size*	Capital Invested	Sales	Overheads	Overheads Net of O.C.
	(In $1,000)		%	%	%	%
1	24 or less	12	84.6	66.7	84.9	93.1
2	25 - 49	37	97.2	73.9	86.7	87.8
3	50 - 99	72	97.2	81.4	92.8	101.6
4	100 - 249	162	97.4	76.4	91.1	97.6
5	250 - 499	353	98.2	76.8	91.4	96.5
6	500 - 999	701	99.9	79.1	95.3	99.9
7	1000 - 2499	1552	101.1	85.2	96.1	98.9
8	2500 - 4999	3487	101.2	90.5	105.5	107.9
9	5000 - 9999	7028	103.0	81.1	94.6	95.9
10	10000 - 49999	20652	104.0	88.4	78.0	77.9
11	50000 - 99999	70480	63.9	70.8	58.7	58.8
12	100000 - or more	434310				

*Size by capital invested; net-deficit and net-income corporations are combined.

The fact that the average firm size by capital invested is about equal in the two groups is not surprising because firms are grouped by their asset sizes. However, size measured by sales is about 20% lower on the average in the net-deficit group. Now if we assume that firms with the same asset size have similar production and financial structures, we may expect that they would incur about the same amount of overhead expenses (net of officers' compensation) irrespective of their actual scales of operation. The last column of the table proves this expectation. Except for the smallest and largest size classes, the ratio of overheads net of officers' compensation turns out to be close to unity.

We have seen that average firms in the net-income and net-deficit groups are not significantly different with respect to capital invested and to overhead net of officers' compensation, but are considerably different with respect to sales. How much of the differences in sales are due to differences in price and output? Also what about the c/p ratio? Columns (1) and (2) in Table 6 below show direct costs as a percentage of sales for each group over all size classes. Let us make two plausible assumptions: (1) The average net-deficit firm sells its product at a price no higher than the average net-income firm of corresponding size does ($p_d \leq p_i$) ; (2) the average net-deficit firm produces its product at average unit direct cost no lower than the aver-

age net-income firm of corresponding size does $(c_d \geq c_i)$.[1]

TABLE 6

Possible Price-Cost-Output Relationships Between the Average Net-Deficit Firm and the Average Net-Income Firm, 1954*

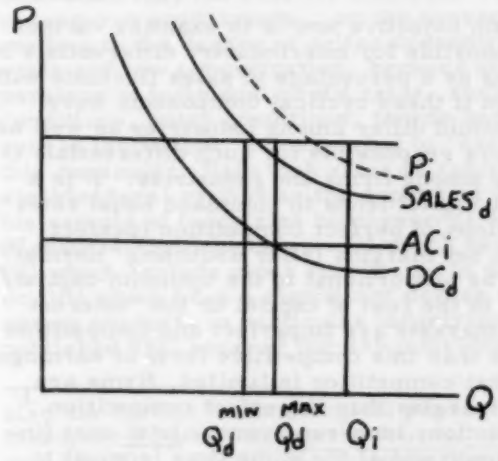
Size Class	c_i/p_i	c_d/p_d	p_d/p_i min.	c_d/c_i max.	Q_d/Q_i max.	Q_d/Q_i min.
1	68.0	74.5	90.2	110.9	74.0	66.7
2	69.2	76.6	90.6	111.4	81.6	73.9
3	70.9	78.1	90.8	111.7	89.6	81.4
4	73.8	79.9	92.4	108.2	82.7	76.4
5	75.3	81.0	93.0	107.5	82.6	76.8
6	75.3	81.9	91.9	108.8	86.1	79.1
7	75.3	84.0	89.6	111.6	95.1	85.2
8	74.5	83.2	89.5	111.5	101.1	90.5
9	72.7	82.1	88.6	112.9	91.5	81.1
10	70.3	86.2	81.5	112.7	108.5	88.4
11 + 12	72.1	86.7	83.1	120.3	85.2	70.8

*Subscripts i and d denote net-income group and net-deficit group respectively.

These assumed conditions set limits to possible ranges for both the price ratio (p_d/p_i) and the cost ratio (c_d/c_i). The column (3) is the ratio of the column (1) and (2), or the ratio of the price ratio to the cost ratio. When the cost ratio is at its assumed minimum, i.e., unity, the price ratio takes its minimum value which is equal to the value shown in column (3). Similarly when the price ratio is at its maximum (=unity), the cost ratio will take the maximum value which is equal to the inverse of column (3) and shown in column (4). When the price ratio is at its maximum, output is at its minimum which is equal to the sales ratio as given in Table 5 and reproduced in column (6) of Table 6. When the cost ratio is at its minimum, output is at its maximum, as given in column (5). These relations are shown in Figure 1 where total cost, sales and direct cost of the net-deficit firm are shown by rectangular hyperbolas.

1. This assumption depends on the particular shape of the marginal cost curves. If the two firms share the same sharply rising marginal cost curve, the net-deficit firm may have a lower average unit direct cost at smaller output. But, then, the net-deficit firm must have a still lower selling price and a fortiori it tends to support the dominance of the second factor (higher competitiveness). There are other complicating factors, such as differentials in wage rates (the net-deficit firm may have lower wage rates) and in materials prices (the net-deficit firm may be paying higher prices for them).

FIGURE 1



A remarkable stability exists in columns (3) and (4). The average net-deficit firm has either 10% lower price or 10% higher unit direct cost than the average net-income firm of corresponding size. Its scale of production is, on the average, 10 to 20% lower. If we strike the middle for a rough estimate, we may state that the average net-deficit firm produced output 15% less at price 5% lower and average unit direct cost 5% higher than the corresponding net-income firm. In view of these figures, we may conclude that the factor which is most responsible for unprofitability of net-deficit firms is absolute shortage of demand in relation to capacity. Inefficiency does not seem to be a crucial factor in this regard--at least among firms of similar size. When demand is relatively short, we may reasonably expect that the firm with a lower demand curve has a more elastic demand upward and less elastic demand downward as compared to the firm with a higher demand curve. Thus, lower output and lower price are probably a more usual combination than lower output and higher direct cost. More detailed data on individual firms are required to substantiate this conclusion of ours. However, existing pieces of information tend to support it.

4. Conclusion. At any given point of time, the distribution of demand among firms is quite important in determing profitability of individual firms. So far as this distribution is stable over time, we can expect that the average percentage gross margin for manufacturing as a whole tends to be stable, too. Changes in demand over the cycle affect all firms alike and thus the relative distribution of demand among firms is expected to remain stable over the cycle. It may be noted, however, that small firms are particularly vulnerable to cyclical changes in demand. Small firms have to share limited markets among themselves with the constant threat of new entry. A large proportion of small firms always has net losses; this proportion is very sensitive to cyclical changes in demand. Unsuccessful small firms go out of business after a very brief life span but new entrants of similar size constantly take their place.[1] Thus, the distribution of demand tends to be stable in spite of the high mortality rates of relatively small firms.

1. For entry rates, mortality rates and life expectancy of the business population, see Betty C. Churchill, "Age and Life Expectancy of Business Firms," Survey of Current Business, XXXV (December 1955), pp. 15-19, and "Size Characteristics of the Business Population," Survey of Current Business, XXXIV (May 1954), pp. 15-24.

3. Interindustry Study

So far we have examined the behavior of profit margins on a more or less global level. In Section 3 we shall turn to more disaggregated observations for a similar analysis.

A. Secular Behavior

1. Problem. Our main objective now is to examine various factors which are believed to be responsible for interindustry differentials in percentage net margins. Net margins as a percentage of sales fluctuate with cyclical changes in demand. But even if these cyclical components were eliminated, percentage net margins would differ among industries as well as among firms. One of the major factors responsible for such differentials is differences in the capital/output ratio among firms and industries. It is a familiar classical proposition that all capital tends to command equal rates of return in the long run under conditions of perfect competition (perfect markets and free entry). Percentage net margins (after deducting "normal" profits for entrepreneurship) tend to be proportional to the optimum capital/output ratios; the proportion is equal to the cost of capital or the "interest rate" plus the profit tax rate. When markets are imperfect and competition is limited, capital tends to earn more than this competitive level of earnings even in the long run. To the extent that competition is limited, firms are able to charge higher percentage net margins than in perfect competition.[1] At the long-run optimal level of production, long-run average total cost (including normal profits) is at its minimum and at the same time is equal to short-run marginal cost and marginal revenue.[2] These are basic and familiar theorems in theory of the firm.

However, let us pose questions of empirical nature such as: To what extent can these theorems explain actual differentials in percentage net margins among firms and among industries? E. g., do rates of return really tend to be equalized in the long run? Is monopoly power strong enough to assure higher rates of return in the long run? If so, by how much? Or does rapid technological progress tend to upset any tendency for even giant corporations to earn their long-run monopoly gains through increasing their rivals' potential as well as actual strength in competition?[3]

2. Previous Studies. These questions can be answered only through careful empirical analysis of actual data. Little empirical work has as yet been made within the given analytical framework. The only previous work along this line is that of Professor Bain.[4] In his 1951 paper Bain pre-

---

1. For a more complete statement of this thesis, see e. g., Arthur Smithies, "The Control of Inflation," Review of Economics and Statistics, XXXIX (August 1957), p. 274.

2. The Lerner degree of monopoly is meaningful only in this long-run equilibrium.

3. See e. g., A. D. H. Kaplan, Big Enterprise in a Competitive System, (Washington, D. C., The Brookings Institution, 1954), p. 244, where he argues that "in few markets can a company trust to its established position to perpetuate profitable margins. The threat of competitive innovation compels it to engage in dynamic competition."

4. Joe S. Bain, "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-1940," Quarterly Journal of Economics, LXV (August 1951), pp. 293-324; Barriers to New Competition, (Cambridge, Mass., Harvard University Press, 1956), especially Chapter 7, pp. 182-204. See

sented the hypothesis that firms in oligopolistic industries of high concentration will tend to earn higher rates of return on the average than firms in less concentrated oligopolies or industries of atomistic structure.[1] He argued that even if the association of high profit rates on sales and high seller concentration may not hold for individual cases, it "should actually hold even in long-run equilibrium...on the average"[2] when we group industries according to the degree of seller concentration into categories of high, moderate and low concentration, because all other factors which cause dispersions of individual profit rates, such as differences in cost and demand conditions, entry conditions, trends and fluctuations in demand, and the rate of innovation, are expected to "average-out" within groups.[3] After this reasoning, Bain took rates of return or profits (after income taxes) on stockholders' equity[4] as the dependent variable. In the actual test, to keep his sample of industries homogeneous and meaningful, he first eliminated (Census 4-digit) industries (i) which have geographically segmented markets, (ii) which exclude close substitutes to outputs included in particular industries, or (iii) which have a significant degree of specialization by principal firms among products in particular industries. Then, he took profit data from a published SEC source[5] for 335 individual manufacturing corporations which

---

also his earlier methodological paper, "The Profit Rate as a Measure of Monopoly Power," Quarterly Journal of Economics, LV (February 1941), pp. 271-293. Franco Modigliani gives an excellent review, with stimulating discussion of Bain's book in his "New Developments on the Oligopoly Front," Journal of Political Economy, LXVI (June 1958), pp. 215-232, which also touches upon the problem we discussed in Section 1A.

Fritz Machlup casts serious doubts on the use of accounting profits and assets for the purpose of measuring the effect of monopoly. See F. Machlup, The Political Economy of Monopoly, (Baltimore, Maryland, The Johns Hopkins Press, 1952), pp. 490-497. Needless to say, considerable adjustments have to be made with accounting profits and assets for a more sophisticated analysis than that undertaken in the present section.

An interesting attempt to measure the effect of monopoly on profits by comparing the percentage gross margin of corresponding U.S. and Canadian manufacturing industries appears in D. Schwartzman, "The Effect of Monopoly on Price," Journal of Political Economy, LXVII (August 1959), pp. 352-362. His analysis is essentially based on his implicit and unsupported assumption that percentage overheads and capital/sales ratios are roughly in the same relation among U.S. industries and Canadian industries.

1. In his more recent work he included conditions of entry in addition to seller concentration to test the hypothesis for 20 selected manufacturing industries.

2. Bain, op. cit., Quarterly Journal of Economics, 1951, p. 295.

3. Ibid., p. 295.

4. Stockholders equity is used instead of sales, because data are more readily available in this form, and on the average any differences in the ratio of equity to sales among industries are expected to average out within groups. Ibid., p. 297.

5. Securities and Exchange Commission, Survey of American Listed Corporations, 1936-1940.

he classified into 42 industries according to the Census of Manufactures classification and then he computed industry-average profit rates over the period 1936-1940. For the independent variable, concentration, Bain took 1935 Census of Manufactures concentration ratios, namely the proportion of the value of product accounted for by the 8 largest firms in Census industries. Taking cognizance of a large number of "random" elements working on rates of accounting profits, he restricted his test to a comparison of group averages of these industry-average profit rates over the period and found "a fairly high average level of profit rates down to the 70 per cent concentration line, a much lower average level down to the 30 per cent line, and (based on a very small sample) a higher level again below the 30 per cent line."[1]

Several objections may be raised against Bain's treatment of data and variables. First, the rate of excess profit on sales is a more relevant measure than the rate of excess profit on equity. Stockholders' equity depends on financial structures of industries (e.g., the debt-equity ratio); the assets/sales ratio is economically more meaningful than the equity/sales ratio.[2] Secondly, his sample overrepresents concentrated industries; one half of the sample belongs to the high concentration group and only one tenth to the low concentration group. Thirdly, each industry is represented by a few number of large SEC-listed corporations. Less concentrated industries require a larger number of firms for fuller representation. It may give a serious bias to the statistical results.[3] Fourthly, though he took considerable pains in screening industries, he did pay little attention to the fact that large multiplant firms ordinarily go beyond the boundaries of census 4-digit industries.[4] Fifthly, though he made additional tests about the effects on profit rates of other variables, such as the durability of goods and character of buyers, he made none with respect to by far more important variables, such as differences in trends and fluctuations in demand and the rate of technological innovation.

3. New Empirical Test. In what follows we report an empirical analysis of interindustry differentials in percentage net margins in manufacturing for the postwar period. It is concerned with the questions raised earlier in this section, among which Bain's hypothesis is included. The

---

1. Bain, op. cit., p. 313. The high concentration group (22 industries) had a group average of 12.1%, the moderate concentration group (16 industries) a group average of 5.7%, and the low concentration group (4 industries) an average of 11.7%. (The last two groups combined had a group average of 6.9%.) Dividing industries into two groups by the 70 per cent concentration line, he finds the .1% level of significance by the z-test of the high concentration-high profit thesis.

2. For these points, see F. Modigliani, op. cit., p. 223.

3. In his later book Bain justified this procedure on the ground that the average cost of dominant firms in each industry will presumably tend to be close to the minimum average cost attainable in that industry and that, therefore, in assessing the level of the limit price relative to the normal level of unit cost, the rate of profits of such firms is a better measure. However, for other purposes, such as comparing performances of individual industries, it is not necessarily a good one.

4. Census industries are based on establishments producing mainly primary products included in particular industries; thus, any error coming from this source is minimized.

multiple regression approach is adopted to test the statistical significance of the effects of various variables on our dependent variable, percentage net margins.

(i) Sample. The main source of data for this study is, as in Section 1, Statistics of Income.[1] Corporations submitting income tax returns are classified into minor industry groups corresponding more or less to the 3-digit level of Standard Industrial Classification. The Manufacturing sector is subdivided into 123 such minor industry groups. We eliminated one industry, "miscellaneous manufacturing (not elsewhere classified)," and eight industries labeled "unallocable" to which are assigned corporations whose main activities cannot be classified any further than 2-digit groupings. Our sample consists of the remaining 114 industries.[2] The underlying industry classification system is considerably more aggregative than 4-digit classification. As we pointed out, however, so long as we use company data, 4-digit classification and sometimes even 3-digit classification are too narrow for some corporations. The latter seems to be a more reasonable one for our purpose.[3] Besides, the large coverage of the manufacturing sector in our sample offsets any disadvantage from the degree of aggregation of the data.

On the basis of the same definitions as in Section 1,[4] we computed percentage net margins, capital/sales ratios, and net-deficit sales ratios for each industry for each year in the period, 1948-1956, except for 1952.[5]

---

1. Basic data were collected from Source Book of Statistics of Income, unpublished tabulations of corporate income tax returns by the Internal Revenue Service of United States Treasury.

2. For the Internal Revenue industry classification in comparison with the Census classification, see Statistical Appendix, Table 1. There are about 450 4-digit manufacturing industries in the Census (or 420 if Internal Revenue "miscellaneous manufacturing, n. e. c." is eliminated).

3. The question of how to define an industry for the purpose of empirical analysis is yet to be solved. However, for our present purpose of measuring the effects of industrial concentration and other factors on interindustry profit differentials, it is sufficient if our industry consists of firms competing more or less directly with each other in the common market.

4. Various adjustments are required to make accounting data conform to our theoretical concepts. However, we encounter more problems than solutions in the process of adjusting accounting data. These problems refer to how to deal with depreciation allowances, officers' compensation, rent paid on business properties, and so on. More serious is the question of valuation of fixed assets. In a period of rising prices, an industry with a higher proportion of fixed assets to total assets would have its assets relatively undervaluated. No adjustments are made for this factor.

5. No tabulations were compiled for the year 1952 by the Internal Revenue Service. The choice of 1948 for the initial year is because the new Standard Industrial Classification was adopted in 1948 and this resulted in some break in classification from previous years. 1956 is the last year for which the data were available at the time this study was completed.

Then, we computed unweighted averages of these ratios for each industry over the period. We use these unweighted averages for individual industries as our basic variables.

(ii) Variables.

(a) Percentage net margins. The net margin as a percentage of sales[1] for our manufacturing industries is taken as the dependent variable which has to be explained by other independent variables.

(b) Capital/sales ratios. The first of these independent variables is apparently the capital/sales ratio.[2] More capital intensive industries are expected to earn higher net profits on sales in the long run. The functional relationship between percentage net margins and capital/sales ratios is expected to be linear. Insofar as the linear function remains unchanged, the use of time-series averages does not affect statistical estimation even if our variables may show some variations over time. Drastic changes in capital/sales ratios over time, however, may imply unstable price-cost relationships. Three industries were eliminated from our sample on this score.[3]

(c) Industrial concentration. Our second independent variable is industrial concentration as a measure of the extent of monopoly power. Concentration ratios for 4-digit manufacturing industries in 1954, prepared by the Bureau of the Census,[4] as measured by the proportion of the

---

1. Incomes from sources other than manufacturing and related activities are not included in total sales. Since our purpose is to compare market performances, other incomes must be excluded. However, should these other incomes be included, rankings of industries by rates of return would be little affected; but they tend to reduce variations of rates of return on assets and thus, if included, would tend to reduce the measured effect of such variables as industrial concentration. Net margins are computed before deducting income taxes. Since net-deficit firms are not taxed, for our present purpose profit margins after income taxes are not comparable among industries. Income tax rates varied to some extent in this period from year to year, but since all profit-making firms are affected alike, their effect on net margins may be considered as universal unless industries are very different in ability to shift taxes. Interest paid is excluded from net margins as a part of overhead expense. Theoretically, it should have been included in net margins rather than in overheads. However, interest paid was only a small fraction of sales (.43% for all manufacturing in 1956) and interindustry differences in this ratio are negligible.

2. Capital invested is defined to exclude holdings of securities. Average capital/sales ratios were adjusted by multiplying them by (1 - net intangibles/capital invested) for the year 1956 so as to eliminate intangible assets from capital invested.

3. Year-to-year fluctuations in the capital/sales ratio is due to the composite effects of secular shifts in real capital/output ratios, changes in relative prices and asset valuations, and fluctuations in demand relative to capacity. In terms of coefficients of variation with respect to capital/sales ratios over the period, 114 industries in our sample are distributed as follows:
0 -4.9% (28), 5.0-9.9% (58), 10.0-14.9% (22), 15.0-19.9% (3), 20.0% or more (3).
Three industries with the extreme variation of capital/sales ratios were eliminated from our sample (Industries 372, 396, and 399).

4. The Bureau of the Census, U.S. Department of Commerce, The Propor-

value of shipments accounted for by the 4 largest companies, are averaged with the industry value of shipments as their weights to form concentration ratios corresponding to our Statistics of Income industries. These ratios cannot be taken for a quantitative measure of "monopoly power"; they are, at best, suggestive of such. Thus, we divide our sample into three groups according to the degree of concentration. They are called the high, moderate and low concentration groups respectively. The high concentration group consists of industries with concentration ratios of 50% or more; the moderate concentration group with concentration ratios between 30% and 49%; and the low concentration group with concentration ratios of 29% or less. The boundary lines are chosen more or less arbitrarily. However, the 50% (4-company) line is roughly equivalent to the 70% (8-company) line which Bain found to be significant in dichotomizing his sample. They are also chosen to make industries distributed not too unevenly among the groups.[1] Two dummy variables, z_1 and z_2, are introduced. They take values (0,0) for the low, (1,0) for the moderate, and (0,1) for the high concentration group. These variables are introduced linearly to our functional relationship. Statistical estimates of the coefficients of the two dummy variables measure the average effects of moderate and high concentration on price-cost gaps vis-a-vis low concentration.[2]

(d) Market segmentation. As the third independent variable, geographical market segmentation is considered. When local and regional markets are sufficiently insulated from each other, national concentration ratios are a misleading measure of "monopoly power"; firms with local or regional market control can be expected to earn higher profits than purely competitive firms in national markets, even if industry concentration ratios are the same. It is illuminating to compare the effects of local-regional monopoly and national monopoly on price-cost relationships. Therefore, another dummy variable, z_3, is introduced. It takes the value 0 for industries with national markets and the value 1 for industries with local-

---

tion of the Shipments (or Employees) of Each Industry or the Shipments of Each Group of Products Accounted for by the Largest Companies as Reported in the 1954 Census of Manufactures, (Washington, D. C., Government Printing Office, 1957). See also Report of the Federal Trade Commission on the Concentration of Productive Facilities, 1947, Total Manufacturing and 26 Selected Industries, (Washington, D. C., Government Printing Office, 1949) for a discussion of practical difficulties in estimating concentration ratios on a company basis. No concentration ratios were available for three Ordnance and Accessories industries; they belong to our high concentration group according to estimates based on Source Book of Statistics of Income. In some industries, corporations are engaged in extensive nonmanufacturing activities, such as mining and distribution. Thus, Census concentration ratios which were based on manufacturing activities alone may be incomparable with Statistics of Income data.

1. The high, moderate and low concentration groups consist of 29, 34, and 51 industries respectively.

2. As a statistical technique it is equivalent to taking group averages for comparison. These groupings are required also because of possible nonlinear effects of the concentration variable on profits, as Bain's findings suggest.

regional markets. Sixteen industries were identified as such.[1] In this connection, it must be mentioned that no account has been taken of international competition which may have a significant influence on the high concentration group in particular.[2] No readily available data were found for our purpose.[3]

(e) Dynamic variables. As we discussed earlier, percentage net margins are affected significantly by cyclical fluctuations and secular trends in demand. The importance of these dynamic factors is well known but never measured. Bain bypassed these complicating factors on the assumption that they would average out within groups. The assumption is reasonable, but it requires a test before it can be accepted. Besides, it is important to estimate how strongly these dynamic factors act on actual percentage net margins. For normative welfare discussions, we must have some quantitative notion of the effect of monopoly power in the long run; it can be approximated only after the disturbing effects of the dynamic elements are eliminated from our statistics.[4] Let us face the technical problem of how we can measure these dynamic factors.

Percentage net margins can be considered as composed of two components--secular and cyclical. The latter must be eliminated to estimate the former. The cyclical component (n_c) is defined as the difference between cyclical components of percentage gross margins (m_c) and overheads (e_c). We argued that m_c and e_c are both influenced by cyclical fluctuations in demand. e_c tended to be more variable than m_c and they tended to move in the opposite direction from each other. It is not unreasonable to assume that n_c tends to fluctuate more or less in proportion to the degree of excess capacity. How do we measure the latter, then? Unfortunately we have no reliable capacity measures for individual industries. Lawrence Klein recommends trend lines which connect production peaks over cycles as the

---

1. Local-regional-national divisions of markets are based on data given in National Resources Committee, The Structure of the American Economy, Part I, (Washington, D.C., Government Printing Office, 1939), Appendix 8, pp. 264-269. When several industries were combined to form our Statistics of Income industries, our industries were identified as local-regional market industries only if more than 50% of their combined value of shipments were sold in local-regional markets. For the list of industries in this group, see Statistical Appendix, Table 1.

2. In this connection, see George J. Stigler, "The Extent and Bases of Monopoly," American Economic Review, XXXII (June 1942, Supplement), pp. 1-22, especially p. 7. For instance, the relatively low profitability of the tinware industry, which is highly concentrated, is probably due to the international competition to which this industry is exposed.

3. We could have possibly used the 1947 interindustry table which lists amounts of competitive imports as well as domestic production unless the year 1947 was subject to abnormal postwar conditions.

4. Arnold C. Harberger attempted to measure misallocation of resources due to monopoly power on the basis of industry rates of return (on net worth), 1924-28 averages. See his "Monopoly and Resource Allocation," American Economic Review, XLV (May 1954, Papers and Proceedings), pp. 77-78. He accepted actual average rates of return without adjustments for dynamic departures of profit rates from their normal long-run level.

approximate measure of industry capacity for ordinary pragmatic uses.[1] This approximation, however, can be followed only when cycles have large upswings. In a period of moderate business boom, such as 1954-56, it is hard to tell whether actual peaks of industry output really reached the capacity limit. Because of this difficulty we had to take another less satisfactory approach. First, we took Federal Reserve Board indexes of production for manufacturing industries on a basis comparable with our industry classification after some recombinations.[2] Fifty-five such series of annual indexes of production were obtained,[3] covering about one half of our sample. Then, we fitted exponential trend lines to the 1947-1956 annual series, computed trend rates of growth and the ratios of actual output to trend output. If the trend line is fairly parallel to the capacity line, large variations of the ratio of actual output to trend output over the given period mean a high degree of excess capacity on the average during the period. Hence, we took standard deviations of the actual/trend output ratios as a measure of average cyclical excesses of capacity over production for individual industries. There is also a possibility of chronic excess capacity in industries whose rates of growth are small or even negative. To take an account of this fact, the trend rates of growth are included as an additional variable.

FRB indexes of production, however, are available only for one half of our sample. To save information contained in the other half, we have to seek for alternative demand-capacity measures. For one such, we computed percentage rates of change in capital invested between 1948 and 1956, which may be compared with trend rates of growth in output.[4]

Another alternative may be suggested by our net-deficit sales ratio. We demonstrated for total manufacturing and total mining its close correlation with fluctuations in output. For individual industries we can reasonably expect the same close correlation between them. Besides, the net-deficit sales ratio reflects not only cyclical fluctuations but secular trends in demand in relation to capacity. However, one serious trouble arises if we try to use this d-ratio for an interindustry comparison. The magnitudes of the ratio depend on the degree of integration within industries as well. In an industry with a few large firms and a periphery of small firms around, the d-ratio will certainly be kept low even though it still shows a certain degree of cyclical fluctuation. On the other hand, an industry with an atomistic structure will show a higher d-ratio for the same degree of excess capacity.[5] The net-deficit sales ratio then is a function of the degree of excess capacity and the concentration ratio. Instead of throwing away this variable, we in-

---

1. L. R. Klein, "Some Theoretical Issues in the Measurement of Capacity," Econometrica, XXVIII (April 1960), pp. 272-286.

2. For the classification system of the Federal Reserve Board indexes of production, see Federal Reserve Bulletin, XXXIX (December 1953), "1953 Revision of the Monthly Index of Industrial Production."

3. One industry (372) was later removed from the 55-industry sample because of its highly variable capital/sales ratio.

4. For the 54-industry sample we obtain a correlation coefficient of 73 per cent between growth rates in output and assets.

5. Because in a large firm net losses in some part of it are added up with net profits in the other part with the result that the net deficit ratio is lower than when this firm is broken up into a number of independent units.

troduced additional simplifying assumptions and made the d-ratio a function of the degree of excess capacity and the two dummy variables for concentration grouping.[1] The statistical procedure is to include d as one of the in-

---

1. There is a certain non-linear relationship between the d-ratio and the concentration ratio as illustrated in Figure 2 below. As the figure suggests, the non-linear curve may be approximated by a step function (like broken lines in the figure). Now let us assume that our three concentration groups coincide with the three steps depicted in the figure. Also assuming that changes in the degree of excess capacity have similar effects on the d-ratio in all groups, we may express the approximated step function as follows:

$$d = a_o y + b_1 z_1 + b_2 z_2 + u ,$$

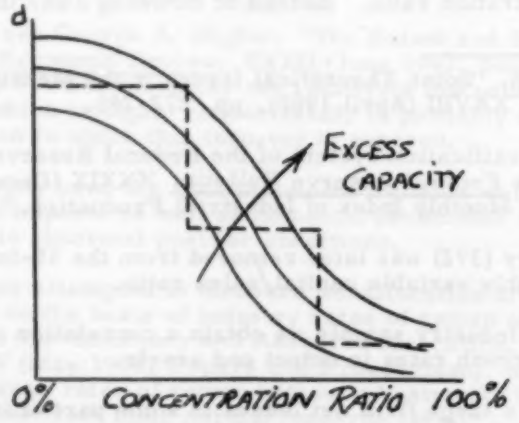
where y stands for the degree of excess capacity and z_1 and z_2 for the two dummy variables for concentration grouping and take values $(0,0)$ for low, $(1,0)$ for moderate, and $(0,1)$ for high concentration.

We assume that this relationship holds approximately in the relevant range of excess capacity. Since we do not know the actual values of y, it is necessary to make another assumption for estimating the coefficients appearing in the equation. Let us take unweighted averages of d and v over the period for each industry and then consider their concentration-group averages, which we designate with subscripts 0, 1, and 2 respectively for low, moderate, and high concentration. Then, we obtain

$$\bar{d}_o = a_o \bar{y}_o + b_o + \bar{u}_o ,$$
$$\bar{d}_1 = a_o \bar{y}_1 + b_o + b_1 + \bar{u}_1 ,$$
$$\bar{d}_2 = a_o \bar{y}_2 + b_o + b_2 + \bar{u}_2 .$$

If we can assume that the concentration groups have similar distributions of dynamic disturbances among their member industries, then the three group averages of y can be considered as being equal to each other and b_1 and b_2 can be estimated as the differences between group averages of d.

FIGURE 2



dependent variables in our regression on percentage net margins and to adjust the concentration coefficients on the basis of the assumed relationship.[1]

(f) Changes in labor productivity. As the last of the independent variables we take changes in labor productivity. They are intended to be a measure of the rate of technological innovation. Industries with rapid technological innovations may be expected to be able to lower costs of production or to introduce new and more differentiated products and thus reap higher rates of return than in the absence of technological innovations. At least this is a popular view on the effect of differential changes in technology on industries' productivity. The present variable is included to test this hypothesis. Labor productivity is measured by output per production-worker manhour in each industry. Census production indexes for 1954 relative to 1947[2] and Census manhour data for 1947 and 1954[3] were taken and recombined to match our industry classification and then labor productivity for 1954 relative to 1947 computed.[4] Changes in labor productivity from 1947 to 1954 cluster between 20% and 40%. Differences in productivity changes within a narrow range may be insignificant as far as our hypothesis is concerned. Besides, there is no a priori reason that large increases in productivity leads to proportional increases in percentage net margins. We grouped our industries accordingly into three categories, namely the high productivity group with 145% or higher 1954/1947 labor productivity, the moderate productivity group with 115 to 144% labor productivity, and the low productivity group with 114% or less labor productivity. The dummy variables z_4 and z_5 take values (1, 0) for the low productivity group, (0, 0) for the moderate productivity group and (0, 1) for the high productivity group.

(g) Other variables. There are other important variables which may affect the level of percentage net margins, as Bain clearly

---

1. To the extent that our assumption of equal average fluctuations of the degree of excess capacity among the concentration groups is not satisfied, our findings must be qualified.

For the 54-industry sample, assuming that the average degree of excess capacity is a function of the trend rate of growth (Q_g) and the standard deviation of the actual/trend output ratio (Q_{sd}), we obtained the following least-squares regression equation:

$$d = \underset{(.131)}{.235}\, Q_{sd} - \underset{(.177)}{.499}\, Q_g - \underset{(1.479)}{3.539}\, z_1 - \underset{(1.528)}{8.367}\, z_2 + 10.730 \quad (\bar{R} = .62)$$

The coefficients for z_1 and z_2 and the constant term should be expected to be about equal to $\bar{d}_1 - \bar{d}_o$, $\bar{d}_2 - \bar{d}_o$ and d_o. The latter are -3.195, -7.399 and 10.878. Our assumption of equal means of the average degree of excess capacity seems to be reasonable, though the high concentration group may have a slightly larger true mean than the statistical average suggests.

2. Bureau of the Census, Department of Commerce, United States Census of Manufacturers, 1954, Volume IV, Indexes of Production, (Washington, D.C., Government Printing Office, 1958).

3. Idem, Vol. I, Summary Statistics, (Washington, D.C., Government Printing Office, 1957).

4. Output indexes were weighted by cross weights based on value added which were supplied in the volume. No comparable output and manhour data were published for three Ordnance and Accessories industries and they were excluded from our sample.

demonstrated in his book, namely conditions of entry. Available data, however, are yet so fragmentary that we have to omit these variables entirely from our analysis.

TABLE 7 (A)

Characteristics of the Samples*

		Sample A		Sample B	
Number of industries		54		110	
Concentration	Low	23		51	
	Moderate	17		33	
	High	14		26	
Local-regional Markets		13		16	
Productivity changes	Low			18	
	Moderate			66	
	High			26	
		Mean	St. Dev.	Mean	St. Dev.
Percentage net margins (in %)		7.19	4.27	6.82	3.88
Capital/sales ratios		57.33	17.18	55.73	14.99
Net-deficit sales ratios		7.95	5.44	9.36	6.46
Rates of growth in assets		7.92	6.85	8.37	6.73
Rates of growth in output		2.47	3.39		
Standard deviations of actual/trend output		6.89	4.91		

*For basic data for individual industries, see Tables 1 and 2, Statistical Appendix.

4. Statistical Result. Table 7(A) presents the characteristics of the two samples; Table 7(B) gives the statistical results of least squares regression. $\bar{R}^2$ ranges between 48% and 60%. In view of crudeness of our data and variables, these results should be thought of as fairly satisfactory. We must, however, keep in mind that we can make no precise quantitative statements and that our conclusions must always be qualified. It is interesting to note that the first regression fares best among the three regressions tried on the 54-industry Sample A. In the second regression, after adjustment for the concentration effects, the concentration coefficients become close to those in the first regression. This gives some support to the d-ratio as a usable measure of excess capacity. The third regression yielded fairly comparable estimates with the other regressions.[1]

(i) Capital/sales ratio. The estimated coefficient for this

1. The constant terms in the first and third regressions are negative, while that in the second is positive. The negative effect of the average d-ratio is separated out in the second regression.

TABLE 7 (B)

Regression Coefficients on Percentage Net Margins (%), 1948-1956

Variables	Expected Sign	Sample A			Sample B	
		I	II	III	II	III
Capital/Sales Ratio (%)	+	.110 (.025)	.099 (.029)	.118 (.026)	.090 (.019)	.140 (.020)
Net-Deficit Sales Ratio (%)	-	——	-.294 (.110)	——	-.370 (.045)	——
Trend Rate of Growth in Output (%)	+	.528 (.118)	——	——	——	——
Standard Deviation of Actual/Trend Output (%)	-	-.059* (.086)	——	——	——	——
Rate of Growth in Assets (%)	+	——	——	.209 (.060)	——	.220 (.042)
Concentration, Moderate	+	-.067* (.988)	-.475* (1.043)	.478* (.974)	-.578* (.604)	1.069* (.709)
Adjusted**			.465* (.815)		.939* (.609)	
Concentration, High	+	3.308 (1.095)	.582* (1.386)	2.586 (1.098)	-1.289* (.673)	1.050* (.725)
Adjusted**			2.759* (1.474)		1.354 (.691)	
Local-Regional Markets	+	2.760 (.993)	1.542* (1.127)	2.145 (1.042)	.973* (.687)	.594* (.781)
Productivity, Low	-	——	——	——	.242* (.675)	.153* (.773)
Productivity, High	+	——	——	——	.365* (.576)	-.069* (.662)
Constant Term		-1.523	3.469	-2.576	5.488	-3.481
R^2		.6229	.5324	.5702	.6300	.5192
$\bar{R}^2$		.5747	.4837	.5254	.6046	.4863
s_u		2.786	3.069	2.943	2.444	2.785
Percentage Net Margin, Sample Mean (%)			7.188		6.822	

Standard errors of estimates are shown in parentheses below coefficients.

* Not significant at the 5% level.

** Adjusted for the possible effect of concentration on the d-ratio. Standard errors are estimated by the usual approximation formula.

variable shows that the average competitive level of rates of return on capital after eliminating cyclical and secular demand factors was somewhere in the neighborhood of 10% before tax. Income taxes cut it down to approximately one half, i.e., 5 to 6%, a fairly comparable figure with long-term interest rates in the period.[1]

(ii) Rates of growth. It is significant that rates of growth had a strong effect on the magnitude of percentage net margins. In all our regressions, the rate-of-growth variable, whether measured in terms of output or assets, has statistically significant coefficients. For the 54-industry sample, the unweighted sample mean and standard deviation of trend rates of growth of output per annum are 2.47% and 3.39%. For industries with rates of growth one standard deviation higher than the average, percentage net margins were higher by 1.79%, or 25 per cent higher than the sample mean (7.2%). In terms of rates of growth in assets, we obtain a similar picture: growth rates one standard deviation higher than the average imply percentage net profit margins 20% higher than the average. For a representative industry with the average capital/sales ratio, which is somewhere around 50 to 55, the normal percentage net margin may be estimated at 5 to 6 per cent from the results of our regressions. This is the normal level which should be obtained in the stationary state where the economy neither grows nor contracts. Actually, during our period, 1947-1956, the average rate of growth per annum was 2.5% if measured in terms of output and about 8% if measured in terms of assets. The results of our regressions imply that this average rate of growth raises the percentage net margin by 1.3 to 1.8 percentage points. The difference is about 25% of the normal level. In other words, our representative industry if growing at the average rate tends to earn 25% more than in the stationary state. These estimates, though rough, would give us some quantitative notion on the interaction between rates of expansion and flows of profits.[2]

(iii) Cyclical fluctuations in demand. In the 54-industry sample the coefficient for Q_{sd} has a right sign but is not significantly different from zero.[3] We can also see that the effect of interindustry differ-

---

1. In all the regressions, the constant terms are not significantly different from zero (in the second regression if we subtract the average effect of net-deficit sales from the constant term). In other words, after eliminating all other possible effects, industries with average cyclical fluctuations and secular trends in demand tended to earn about equal rates of return, irrespective of their capital intensity.

2. However, there is a small degree of positive correlation between rates of growth and fluctuations of output around its trend line. For our sample we obtained a positive correlation coefficient, 14 per cent.

3. It may be due to the fact that our capital/sales ratios are negatively correlated with actual/trend output ratios. For the 54-industry sample, coefficients of correlation between these ratios are distributed as follows:

-1.00 to -.61 (18), -.60 to -.21 (16), -.20 to .19 (13), .20 to .59 (6), .60 to 1.00 (1).

The distribution suggests that capital/sales ratios are influenced probably more strongly by cyclical fluctuations in demand than any other factors. However, we cannot estimate normal values of capital/sales ratios without knowing the average degrees of excess capacity.

ences in cyclical fluctuations in demand as measured by the standard deviations of the actual/trend output ratio on the d-ratio is not strong. This is partly due to the fact that there probably exists a considerable high minimum level of the d-ratio for the low concentration group even at near capacity production and partly due to the inadequacy of our measure. In any case, on the average over the business cycle, percentage net margins tend to be substantially lower than their normal long-run level in the absence of such cyclical fluctuations.[1]

(iv) Industrial concentration. Now we come to our concentration thesis. For the 54-industry sample we found that the high concentration group had a significantly positive effect on percentage net margins, but the moderate concentration group did not. This finding tallies with Bain's previous conclusion. However, with the larger sample statistical evidence is not so clear in this respect. Therefore, it is risky to make quantitative guesses about the effect of concentration. Nevertheless, some quantitative notion is useful. We can probably state as a rough estimate that the high concentration group charged, on the average percentage net margins 2 to 3 percentage points higher than the competitive level. If we compare this differential with the sample mean of percentage net margins, 7 per cent, we can conclude that it is quite a significant factor in income distribution over a period of years.[2] Thus, the effect of high seller concentration may be quite strong in determining price-cost relationships among industries. In contrast, the moderate concentration group did not have any significant influence on price-cost gaps as compared with the low concentration group.[3]

(v) Local-regional markets. The effect of geographical market segmentation is estimated to be fairly strong in the 54-industry sample, but is reduced to a smaller fraction, positive but not significantly different from zero, in our larger sample.[4] Considering the fact, it seems

---

1. The group mean of the d-ratio for the low concentration group is 12.3 per cent for the 110-industry sample. The net effect of the dynamic factors may be measured at 4 to 5 per cent in comparison to the actual sample mean of percentage net margins, i.e., 7 per cent. Thus, in the absence of these dynamic factors, average percentage net margins may have been about 11 to 12 per cent.

2. Our net margins are net of interest paid and depreciation, amortization and depletion allowances. To the extent that highly concentrated industries have faster rates of depreciation and are more dependent on borrowed funds, they would have larger percentage net margins gross of these payments. The effect of concentration, then, should be higher. Similarly, our capital invested is net of depreciation reserves. If we had taken gross assets, there may have been some differences in statistical results.

As for the capital/sales ratio, the group average for the low concentration group was slightly lower than that for the other groups--Low 51, Moderate 62, and High 60 per cent. They are, however, not significantly different from each other.

3. All our conclusions are also subject to our particular groupings of industries into the three concentration categories.

4. In the 54-industry sample, more than a half of the competitive industries were also in the local-regional market group. Thus, the 54-industry sample was biased heavily toward the more concentrated industries. This may account for differences in statistical estimates.

to be safe to conclude that on the average industries with local or regional market control tended to earn a little more than the competitive national industries, but their ability to charge higher prices was less strong than industries in the high concentration group.

(vi) Changes in labor productivity. This variable was included to test the hypothesis that industries will benefit from faster rates of technological progress. As it stands, the coefficients for the two technology dummy variables turned out to be not significantly different from zero. Differential changes in productivity did not seem to have had any noticeable effect on industries' relative profitability. [1]

How much of the observed variance of percentage net margins among industries can be explained by these variables? In the first regression of Sample A, 62% of the total variance was explained by six independent variables; 62% is divided among 4 of these 6 variables as follows: capital/sales ratios 27.5%, growth rates 19.1%, high concentration 9.5%, and local-regional market 6.5%. Output deviations and moderate concentration made negligible contributions.

In the Regression II of Sample B, 63% of the total variance is explained by 7 variables. 63% is divided between capital/sales ratios (20.8%) and the d-ratio (44.4%). Concentration and market variables as well as productivity variables helped little in reducing variances. If the gross effect of the d-ratio is considered as a combined effect of growth, fluctuations, and concentration, we may conclude that the strength of profit rates tending to be equalized among industries is not high. The dynamic and market forces outweigh whatever tendency there is.

5. Conclusion. We have set out to see what factors can account for differences in percentage net margins among industries--in particular, to examine the short-run influence of the familiar classical thesis that profit rates tend to be equalized in the economy and to investigate the effects of monopoly power on price formation. Our statistical investigation, though imperfect in many respects, has revealed that the classical thesis can explain only 20 to 30% of actual variances of percentage net margins among industries, while the concentration factors may explain another 20% and the dynamic factors 20%. The level of percentage net margins which is expected to be normal seems to be compatible with the prevailing long-term rate of interest. The effects of the concentration factors and dynamic factors are to make some industries earn significantly above this normal level of profits. In particular, the dynamic factors can have a strong influence on price-cost structure. Therefore, it is quite difficult to base any normative judgement on actual market performances such as profit rates or price-cost ratios.

B. Cyclical Behavior

1. Problem. In this section our interest shifts from secular to cyclical behavior of price-cost relationships in manufacturing industries. We demonstrated earlier that different sectors of the economy are characterized by different patterns of price-cost behavior in the short run as evidenced by cyclical movements of percentage gross margins in relation to those of per-

---

1. There is a certain intercorrelation between changes in labor productivity and industrial concentration. The correlation is mostly due to a close association of low concentration and low productivity rises.

Technological changes also alter capital/output ratios. This factor is too closely intertwined with changes in assets/sales ratios to allow us any meaningful analysis of our data.

centage overheads and that they may be explained for the most part by differences in their marginal cost curves. Are there any similar differences in cyclical price-cost behavior among manufacturing industries themselves? If so, what factors can be found to explain such differences? Are differences in industrial concentration important in this respect?

Why are differences in cyclical and short-run price-cost behavior important? Do they have any practical significance? To answer this question, it is sufficient to refer to the controversy over causes of price inflexibility in the depression of the 30's. Gardiner C. Means asserted that rigid prices in the economy--mostly of manufactured commodities--were due to growing monopoly power of big businesses which administered prices. He presented as his evidence relative inflexibility of prices of manufactured goods in amplitudes, frequencies, etc., of their changes as contrasted to wide fluctuations of agricultural prices. It is not necessary to go into the details of this controversy here.[1]

2. Previous Studies. Proponents of the price inflexibility thesis usually divide the actual market types into two archetypes--the purely competitive market and the oligopolistic market.[2] In the purely competitive market, e.g., agriculture, prices are promptly adjusted to changes in demand; individual producers have little control on market prices. On the other hand, in the oligopolistic market, changes in demand will be first met by corresponding changes in output; the few firms which control the market do not adapt their prices immediately to changes in demand conditions. They do not maximize short-run profits but aim to maximize their long-run profits. Prices are changed only by discrete steps. The proponents of the price inflexibility thesis believe that in industries with atomistic structures prices will react to changes in demand very rapidly just as in agriculture, whereas prices in oligopolistic markets will be administered by dominant firms so that they become rigid, especially in cyclical downswings. Price behavior in the short run is thus considered to be primarily governed by market structure with little reference to cost conditions.

There is little doubt that in agriculture prices, if without government intervention, behave exactly as in the purely competitive model. But, as we argued earlier, this characteristic should be attributed to the inelastic supply of agricultural products in the short run rather than its market structure. It is unwarranted to apply a pure analogy from the short-run market performances of agriculture to manufacturing industries with different production structures involving different (shorter) time lags for adaptation to changes in demand, even if they are highly competitive.[3]

---

1. For a fuller discussion of this controversy, see R. Ruggles, op. cit., especially pp. 441-464. Also see a review of literature on administered prices by Martin Bailey, "Administered Prices in the American Economy," in The Relationship of Prices to Economic Stability and Growth, Compendium of Papers Submitted by Panelists Appearing Before the Joint Economic Committee, (Washington, D. C., Government Printing Office, 1958), pp. 89-105.

2. See for a recent presentation of this type of argument John M. Blair, "Administered Prices: A Phenomenon in Search of a Theory," American Economic Review, XLIX, Papers and Proceedings (May 1959), pp. 431-450.

3. J. K. Galbriath applied the same dichotomy to examine the efficacy of monetary and fiscal policy on controlling business cycles. See his "Market Structure and Stabilization Policy," Review of Economics and Statistics, XXXIX (May 1957), pp. 124-133. He identified the purely competitive

The proponents of the price inflexibility thesis may be expecting that the short-run price behavior of the two extreme types of markets result in such different patterns of cyclical price-cost behavior as those among different sectors of the economy. Then, in the atomistic markets, cyclical components of percentage gross margins are expected to be very sensitive to changes in demand and thus move counter to movements of cyclical components of percentage overheads. On the other hand, in the oligopolistic markets, cyclical components of percentage gross margins may even tend to change in the same direction as cyclical components of percentage overheads if oligopolistic firms can really change their prices so as to offset declining demand. If so, we should expect in a depression such as the years following 1929 that cyclical components of percentage gross margins would have fallen very sharply in highly competitive manufacturing industries, just as in agriculture, and have remained stable or even risen in oligopolistic industries. The short-run movements of cyclical components of percentage gross margins over such years among industries are, then, closely related to the degrees of market competitiveness.

This logic was followed essentially by Alfred C. Neal who tested G. C. Means' monopoly power-rigid price thesis in a similar interpretation.[1] Rather than working with percentage gross margins, Neal took a number of 4-digit manufacturing industries and compared changes in their absolute gross margins per unit of output over a pair of two years (1929-31 and 1929-33) with their 1935 Census concentration ratios. He found that concentration did have a small but significant influence upon the changes in gross margins, which tended to decline least where concentration was high, though "differential price behavior among industries for both comparisons (1929-31 and 1929-33) is to be explained for the most part by differential unit direct cost behavior rather than by concentration."[2] However, because of his particular choice of changes in absolute unit gross margins for his comparison, his finding about the significant correlation of gross margins with industrial concentration is not conclusive for the cyclical behavior of percentage gross margins.[3] Neal's study, which has been frequently quoted in discussions of

---

markets with "agriculture, bituminous coal mining, forest products, and the staple branches of the cotton textile industry." According to him, in these industries "prices are vulnerable to any reduction in aggregate demand" and "there will be immediate adaptation through reduced prices" to a curtailment of demand (p. 131).

1. Alfred C. Neal, Industrial Concentration and Price Inflexibility, (Washington, D. C., American Council on Public Affairs, 1942).

2. Ibid., p. 165.

3. Should he have used percentage gross margins rather than absolute gross margins (per unit of output) for his comparison, he would probably have found little correlation between concentration ratios and changes in gross margins. According to his definition, his dependent variable, i. e., changes in unit gross margins (with price indexes for 1929 at 100) from 1929 to 1931 or 1933, reduces to the following expression:

$$\frac{\Delta m_{t+1}}{1+m_t} - \frac{\Delta c_{t+1}}{c_t} \frac{m_{t+1}}{1+m_t} ,$$

short-run price-cost behavior, seems to be the most extensive work in this field. The meagreness of analytical, empirical studies in the field attests the difficulty involved in making rigorous analysis of short-run price-cost behavior and makes further work really desirable.

3. Empirical Examination. We attempted to examine short-run changes in price-cost relationships in manufacturing industries with a hope of adding something to our meagre empirical knowledge on this subject. Essentially, we made a test of the hypothesis on cyclical behavior of the percentage gross margin which should be expected from the application of the dichotomy of market types to manufacturing. The result, though highly tentative, is reported below.

We should expect from the hypothesis that over the cycle cyclical components of percentage gross margin and overhead, m_c and e_c, tend to move oppositely from each other in atomistic industries, whereas countermovements of m_c and e_c are less pronounced in oligopolistic industries. Actually for manufacturing as a whole we found that estimated cyclical components of m and e had a negative correlation of -.52 over 1947-1956. Then, if the hypothesis be true, we should expect that such coefficients of correlation, if computed for individual industries over the same period, will take significantly larger negative values (in absolute terms) for a group of competitive industries than for a group of oligopolistic industries. We tested the hypothesis in this form.

The first procedure was to estimate linear trends for both m and e over the period 1948-1956 and to take residuals as estimates of cyclical components of m and e. Then, coefficients of correlation between the latter two were computed for all industries in our sample which we introduced in the last section.

It was found that a serious difficulty was involved in this procedure; linear trends do not apply so well to individual industries as to manufacturing as a whole. One cause for the departure from linear trends is classificatory shifts of firms among industries from time to time. Estimated m_c and e_c are subject to the same "random" disturbances. Inclusion of these random terms in estimated m_c and e_c tended to bias our correlation coefficients upward. The extent of their biasing influence is difficult to guess. However, so far as their influence on individual correlation coefficients remains random and moderate, true differences in distributions of these correlation coefficients within concentration groups should still remain significant, if the hypothesis be true. If we find that these distributions are not significantly different from each other, we should suspect that the hypothesis is not true. The statistical result, presented below, suggests the latter conclusion.

Table 8 summarizes our statistical result by classifying industries according to the size of correlation coefficients between estimated m_c and e_c and industrial concentration ratios.

---

where m is a percentage gross margin and c average unit direct cost as percentage of the 1929 price index. From this it follows that his negative correlation between changes in the unit gross margin and the concentration ratio does not necessarily imply a negative correlation between changes in the percentage gross margin and the concentration ratio. It is quite probable that his correlation was predominated by differential behavior of unit direct cost among industries. For instance, from 1929 to 1931, unit direct cost fell only 5% in the iron and steel industry group, even rose by 5% in the transportation equipment group, but fell by 28% in both the food processing and textile and apparel industry groups--the more competitive segments of the manufacturing sector (see R. Ruggles, op. cit., p. 483).

TABLE 8

Two-way Distribution of Industries by $r(m_c, e_c)$ and Concentration, 1948-1956[1]

$r(m_c, e_c)$	Concentration				Total
	Low	Moderate	High	L-R market	
+ .70 to +1.00	7	6	7	6	26
	(10)	(6)	(6.5)	(6)	
+ .20 to + .69	12	6	7	3	28
	(10)	(7)	(7)	(4)	
- .19 to + .19	8	8	7	3	26
	(10)	(6)	(6.5)	(4)	
- .69 to - .20	10	5	6	3	24
	4	1	0	1	5
-1.00 to - .70	(11)	(7)	(7)	(4)	
Total	41	26	27	16	110

1. Four industries are omitted because of their extreme cyclical fluctuations. Numbers in parentheses are those expected when there is no intercorrelation between the two variables.

As it shows, correlation coefficients are considerably biased upwards, though there are still enough variations left. The application of the contingency table test demonstrates that distributions of industries by size of correlation coefficients are not significantly different among concentration groups ($\chi^2 = 5.05 < \chi^2_{.05} = 16.92$). Thus, the statistical result, though admittedly tentative, may be used as an evidence for rejecting the hypothesis.

It is probable that by examining individual price-cost behavior in more detail we may find stronger evidence. For one such examination we turned to inspect data of the early 1930's. After examining percentage changes in unit costs, output and prices for major manufacturing industry groups for 1929-1931, Ruggles concluded that "there does not seem to be any single simple explanation that can account for these divergences" between price drops and unit direct cost drops of individual groups.[1] Some groups had larger price drops than unit cost drops, while others had smaller price drops than cost drops from 1929 to 1931.

On the basis of Statistics of Income data, we computed changes in the percentage gross margin for eleven major manufacturing groups for the same period.[2] For manufacturing as a whole the percentage gross margin showed

1. R. Ruggles, op. cit., p. 483.

2. For the underlying data, see Statistical Appendix of my dissertation.

a small fall from 1929 to 1930 and then leveled off to 1931. When we move to more disaggregated data, we find that m did decrease from 1929 to 1930 for all groups but one--Food, Beverage and Tobacco combined. In the last group m increased from 18.7% to 20.2%. From 1930 to 1931 five groups including the Food group increased their percentage gross margins (Textile-Apparel, Leather, Rubber, Chemicals including Petroleum Products), while the other six had continually declining percentage gross margins (Lumber-Furniture, Paper, Printing and Publishing, Stone, Metals). It is of interest to note that four of the first five groups are consumer non-durable industries. In the Food group, demand was cyclically least flexible and the recovery of output began in 1932, a year ahead of general recovery in 1933. In the other three consumer-goods industries sharp contractions of demand from 1929 to 1930 slowed down in 1931. The Chemicals group was an only exception. In the last six groups percentage gross margins continued to decline from 1929 to 1931. Four of them are durable-goods industries, whose demand continued to contract from 1929 to 1931 with no slackening in pace. Demand for products of the Paper and Printing-Publishing groups behaved similarly over the period.

Generalizations are always hazardous, but in this case it seems that cyclical flexibility of demand was closely related to the downturn and upturn of the percentage gross margin. If we compare the 1929-1931 period with the postwar period, a peculiar coincidence shows itself up. A majority of industries in the Food-Beverage-Tobacco group had very high correlation coefficients between estimated m_c and e_c for the period 1948-1956; in these industries percentage gross margins tended to move very closely with percentage overheads over the cycles. The close association between them does not seem to be due to defects in data. For the sake of completeness, we classified industries according to their 1929-31 behavior of the percentage gross margin and tabulated them in Table 9 below.

TABLE 9

Two-way Distribution of Industries by $r(m_c, e_c)$ and Industry Groups, 1948-1956

Industry Group	Food, Beverage, Tobacco	Textile-Apparel, Rubber, Leather, Chemicals, Petroleum-Coal Products	Lumber-Furniture, Paper, Printing, Stone, Metals, Misc. Manufactures	Total
Changes in m, 1929-31	++	-+	--	
$r(m_c, e_c)$				
+ .70 to +1.00	12	4	10	26
+ .20 to + .69	1	11	16	28
- .19 to + .19	2	11	13	26
- .69 to - .20	0	4	11	24
-1.00 to - .70	0	1	5	6
Total	15	31	64	110

It is apparent that the Food-Beverage-Tobacco group did behave quite differently from all other groups combined. The result cannot be explained by a simple coincidence. Admittedly cyclical behavior of price-cost relationships is a complicated matter. It is extremely difficult to explain it fully in terms of other economic variables. They are influenced simultaneously by a host of diverse factors and we may never be able to disentangle them from such aggregative data as we have here relied on. However, our tentative investigation suggests that we should look for such factors not in industrial concentration and a naive dichotomy of market types, but rather in cyclical variability of demand and producers' adaptation of their price-output relationships.

4. Conclusion. We have made statistical examinations of the hypothesis that cyclical price-cost behavior is somehow or other related to market structure or industrial concentration. We can draw a conclusion, though of tentative nature, that short-run movements of the percentage gross margin vis-a-vis the percentage overhead do not seem to lend support to this hypothesis as far as manufacturing industries are concerned.

Appendix: The Interindustry Profit and Wage Differentials

1. Problem. We discussed, in Section 3, interindustry differences in price-cost relationships. Our primary interest was in the price-cost gap rather than in price behavior and cost behavior themselves. Changes in unit cost of production are governed by technological changes which alter factor proportions and by changes in market prices of input factors. Prices of raw and intermediate materials which one industry purchases from others are determined by particular price-cost relationships and cost behavior in other industries which supply them. The purchasing industries have little control of their prices except in strong cases of monopsony and oligopsony. These input prices may be assumed to be fairly flexible and responsive to changes in cost conditions and/or demand conditions.

There, however, is one factor of production which does not conform very well to this general pattern. It is labor. Imperfection of the labor market is well recognized. Wage rates are believed to be influenced by institutional characteristics of employing industries as well as those of employees. Prices of labor are not determined by purely impersonal market forces of demand and cost alone. They are under the influence of a host of institutional and dynamic factors. In discussing intertemporal variations in the interindustry wage structure, previous writers emphasized the importance of differential effects of changes in labor productivity, competitiveness of product markets, profitability of industries, unionism and unionization, and growth of industries. So far as competitiveness of product markets and profitability of industries have differential influences on variations in the wage structure over time, we must expect a certain degree of interdependence between price behavior and cost behavior among industries. In this appendix we shall examine this interdependence.

2. Previous Studies. In the past decade a number of empirical studies were published which examined the differential impact of these variables on variations in the interindustry wage structure.[1] Some of them, Dunlop and Garbarino among others,[2] stressed the potential importance of

1. For a detailed reference, see e.g., Clark Kerr, "Wage Relationships--the Comparative Impact of Market and Power Forces," in J. T. Dunlop, ed., The Theory of Wage Determination, (London, Macmillan, 1947), pp. 173-193.

2. See John T. Dunlop, "Productivity and the Wage Structure," in Income

changes in labor productivity in influencing changes in the wage structure. Starting with a simple premise that differential changes in labor productivity --measured in terms of output per manhour--would lead to differential decreases in unit cost, which must be passed on either to consumers as reduced prices (or improved product quality), to producers as higher profits, or to labor as increased wages, they argued that differential changes in labor productivity would influence intertemporal variations in the interindustry wage structure, the degree of the influence depending on competitiveness of product markets which would control the firms' ability to keep prices from falling so that any gains in productivity may be left to be split between labor and management, and on the degree of union organization which would determine labor's strength in capturing productivity gains. Their statistical tests, based on pre-World War II data, emphasized the importance of differential movements in productivity and differences in industrial concentration among industries in influencing variations in the interindustry wage structure,[1] while the importance of unionism in this connection was not conclusively established.[2]

The basic mechanism envisaged by these writers is that the gains in productivity have to be passed on to consumers in competitive industries, whereas they can be retained and shared in some proportion between workers and employers in oligopolistic industries.[3] Such a mechanism is believed to describe actual differential variations in the interindustry wage structure fairly well.[4] High productivity increases, usually coupled with expanding output and employment, tend to make industries more profitable and to put them in a position to afford a generous wage policy. These prosperous industries set the pattern of wage increases which will be followed by other less prosperous industries.

In these discussions changes in productivity are given a dynamic role in influencing variations in the interindustry wage structure. Gains in pro-

---

Employment and Public Policy, (New York, W. W. Norton and Co., 1948), pp. 341-362; Joseph Garbarino, "A Theory of Interindustry Wage Structure Variation," Quarterly Journal of Economics, LXIV (May 1950), pp. 282-305.

1. Garbarino, op. cit., pp. 298-299.

2. Contrary to this conclusion, A. M. Ross and W. Goldner concluded that "unionization is a source of wage advantage." See their "Forces Affecting the Interindustry Wage Structure," Quarterly Journal of Economics, LXIV (May 1950), pp. 254-281. The article discussed, in addition to the effect of unionization, the original level of earnings, the proportion of labor cost in total value of product, concentration, and employment changes.

3. Later studies challenged the validity of the association between productivity differentials and interindustry wage differentials. See e.g., F. Myers and R. L. Bowlby, "The Interindustry Wage Structure and Productivity," Industrial and Labor Relations Review, VII (October 1953), pp. 92-102, and Doris Eiseman, "Interindustry Wage Changes, 1939-1947," Review of Economics and Statistics, XXXVIII (November 1956), pp. 445-448.

4. For a more detailed description of this mechanism, see Lloyd G. Reynolds, "Wage Behavior and Inflation: An International View," in Wages, Prices, Profits and Productivity, (New York, The American Assembly, Columbia University, 1959), pp. 109-136, especially pp. 120-122 and pp. 126-127.

ductivity are divided among consumers, producers and employees. Their relative shares are considered as determined primarily by the degree of product market imperfection and only to a minor extent by the degree of union organization. We, however, found that there is some doubt on the thesis that interindustry differentials in profitability are influenced by differential changes in productivity. A question may be raised that, if our previous finding is true, the mechanism described above may not be a true description of interindustry wage behavior. The statistical examination of this question is undertaken in the rest of the present appendix.

3. Empirical Examination.

(i) Sample. Our basic inquiry centers on the productivity-competitiveness-profitability-wage variation thesis. To perform statistical tests on this hypothesis, we made use of our industry rates of return as a measure of the profitability of industries. To match our industry classification system, we combined Bureau of Labor Statistics industry data on average hourly earnings and employment of production workers into comparable groupings.[1] We obtained wage and employment data for 77 such manufacturing industries for 1947 through 1957 and 99 for 1951 through 1957.[2] Other data were already introduced in this study.

(ii) Statistical approach. In carrying our statistical tests on variations in the interindustry wage structure, there is a technical difficulty. The interindustry wage relationships remain very stable over a considerable length of time.[3] Average hourly earnings in different industries tend to move closely together from year to year, reflecting the effect of aggregative fluctuations in demand and of the imitative process of wage advances. Changes in average hourly earnings contain a portion which is due to this overall effect. If we take percentage changes of average hourly earnings (as Dunlop and Garbarino followed) over a period of years, we assume implicitly that the overall factors tended to alter interindustry wage levels in proportion to their original levels of earnings and that factors influencing variations within the wage structure had differential effects which can be correctly measured also by percentage changes from the original levels. Considering the

---

1. Wage and employment data were collected from Bureau of Labor Statistics, Individual Historical Summary Tables on Employment and Earnings. Average hourly earnings are still the best available, even though imperfect, measure of wage levels.

2. For the list of industries included in the sample, see Statistical Appendix. Four industries were combined into two due to the unavailability of separate figures. It is possible that average hourly earnings in our sample change because of shifts in the industry mix within our industries. In most cases, however, average hourly earnings in 4-digit industries within larger industry groups were closer to each other within groups than among groups. The errors due to aggregation are believed not to be important. Our statistical test reported below is based on the smaller 77-industry sample. The conclusions, however, do not seem to be affected by this choice.

3. Its stability was discussed by Sumner Slichter, "Notes on the Structure of Wages," Review of Economics and Statistics, XXXII (February 1950), pp. 80-91. The long-run comparison on a larger sample was made by Donald Cullen, "The Interindustry Wage Structure, 1899-1950," American Economic Review, XLVI (June 1956), pp. 353-369.

importance of the original levels of earnings, Ross and Goldner took absolute changes in average hourly earnings.[1] Their choice presupposes the prevalence of the across-the-board type of wage advances so that workers in different industries tended to obtain the same constant absolute changes in wages which may be ascribed to the overall effect. Depending on which definition we follow, we may arrive at entirely different interpretations of the same facts. Suppose, for instance, over the particular period, average hourly earnings in different industries tended to increase by the same percentage from their original levels. According to the first definition, this movement shows up as no difference in interindustry wage variations; whereas by the second definition, low-wage industries have smaller increases in wages than high-wage industries after allowing for the overall effect. If we are to test the effect of a variable, say industrial concentration, on the interindustry wage variations, the absolute-change approach may support the hypothesis, while the relative-change approach will reject it.

Actually over our period, 1947-1957, average hourly earnings in our sample of industries tended to increase by more or less constant absolute amounts from 1947 to 1952. The interindustry wage structure thus tended to narrow considerably. But from 1952 to 1957 low-wage industries tended to lag behind the pace of wage advances in high-wage industries. Thus, the wage structure was widened again.[2] The divergence of wage behavior over

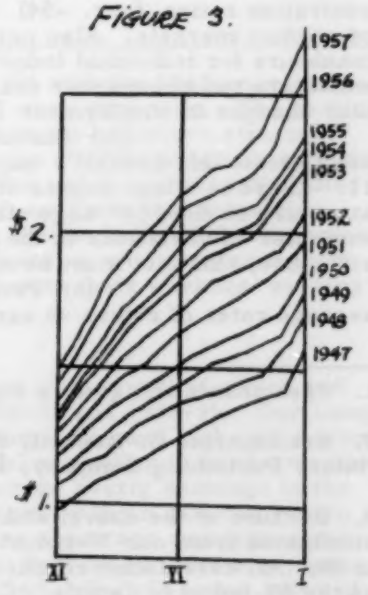
1. See Ross and Goldner, op. cit., p. 257 ff.

2. Figure 3 shows movements of average hourly earnings in our sample of 77 industries during 1947-1957. These industries are divided into 11 ranges by the level of average hourly earnings for each year and range medians are plotted in the figure. The upper half of the sample continued to rise from 1947 through 1957, whereas the lower half kept pace till 1952 but then began to lag. See the summary table below in which the differences between medians of the upper extreme range (I), the middle range (VI), and the lower extreme range (XI) are shown.

TABLE 10

Year	VI	I-VI	VI-XI
1947	123	34	28
1948	135	32	34
1949	141	34	35
1950	147	35	34
1951	159	38	40
1952	167	42	45
1953	178	40	51
1954	184	40	57
1955	190	47	62
1956	200	52	61
1957	212	51	70

FIGURE 3.



the post-war period suggests that we would obtain different statistical results from the two approaches depending on choices of years for comparison. The results would be of doubtful quality. To avoid this technical difficulty, we adopted the partial rank correlation approach.

The interindustry wage structure tends to remain stable over a period of years. To begin with, the composition of employment by skill, occupation and sex within individual industries is stable. Aggregative fluctuations in employment, changes in cost of living, the leader-follower pattern of wage increases, and downward inflexibility of wages all contribute to the short-run stability of the wage structure. Indeed, the original levels of earnings are the most important determinant of the wage structure at the end of the period. For our 77-industry sample, the coefficient of rank correlation (Kendall's tau) between the 1947 and the 1957 wage structures is .75.[1]

If the wage structure changed very little over a period, we may conclude that these "binding forces" were strong enough to preclude any influence of other variables which are expected to have differential effects on the interindustry wage structure. The latter effect, then, should be measured by the degree of change in the wage structure from its original position. The influence of the original wage structure must be eliminated to evaluate the importance of other variables influencing interindustry wage variations. Partial rank correlation is the technique which we should follow.[2]

(iii) Variables. For the actual test we ranked average hourly earnings of 77 industries for 1947 (abbreviated as w-47) and for 1957 (w-57). To measure the effect of relative profitability of industries on wage variations, we computed unweighted averages of rates of return (before tax) on capital invested during the period, 1948-1956, for individual industries ($\bar{r}$) and ranked them similarly. They are regarded as having influenced the 1947 wage structure to get at the 1957 wage structure. To test for the effect of labor productivity changes, we ranked output per manhour, 1954 relative to 1947 (Q/L, 54/47), for individual industries. (Correspondingly, we selected the 1955 wage structure for the comparison.) In addition, 1954 concentration ratios (c.r. -54) were ranked as a measure of competitiveness of product markets. Also percentage changes in total production-worker manhours for individual industries were computed between 1948 and 1956, which are two prosperous years in our period, and ranked to measure secular changes in employment (m.h., 56/48).[3]

(iv) Statistical test. Simple and partial rank correlation coefficients (all Kendall's tau) among our variables are presented in Table 11. There is a high degree of dependence of the 1957 interindustry wage structure on the 1947 wage structure. In evaluating the effects of any other variables on variations in the 1957 relative wage structure from the 1947 structure, this fact must be taken into account.

(a) Profitability. Let us first examine the effect of average rates of return on capital. The interindustry rate-of-return structure

---

1. Spearman's rho gives a higher coefficient of .90.

2. See Maurice G. Kendall, Rank Correlation Methods, 2nd ed., (New York, Hafner Publishing Company, 1955).

3. Because of the unavailability of employment figures, 8 industries were eliminated from our 77-industry sample, reducing the number of industries to 69. All correlation coefficients involving m.h., 56/48, were computed on the 69-industry sample. Correlation coefficients between w-47 and w-57 and between w-57 and $\bar{r}$ are .7951 and .3818 respectively.

TABLE 11

Simple and Partial Rank Correlation Coefficients Among the Interindustry Wage Structure and Related Variables, 1947-1957

	m.h. 56/48	w-55	w-57	$\bar{r}$ 48-56	c.r. -54	c.r. +w-47	Q/L 54/47
Simple Correlation							
w-47	.3683	.7990	.7520	.2212	.1525≠		-.0440*
w-55							.0270*
w-57	.4071			.3014	.2348	.6131	
$\bar{r}$	.3987				.1857≠	.2299	.1008*
c.r.-54							.2011
w-47 +m.h., 56/48			.6700	.4025			
Partial Correlation							
w-57(w-47)	.2007			.2101	.1837≠		
w-55(w-47)							.1044*
w-57(w-47+c.r.)				.2086			
$\bar{r}$ (c.r.)							.0659*
w-57(w-47+m.h.)				.1651≠			

* Not significant at the one per cent level.
≠ Not significant at the one per cent level but significant at the five per cent level.
(The standard error for the population with tau = 0 is .0777.)

has a positive coefficient of rank correlation (.22) with the 1947 wage structure and a higher coefficient (.30) with the 1957 wage structure. The partial correlation coefficient of rank order between the rate-of-return structure and the 1957 wage structure net of the effect of the 1947 wage structure turns out to be .21. The coefficient is not very high but yet is significantly different from zero at the 1% level of significance. It suggests that industries which had rates of return above average during the period tended to improve their employees' earning position, whereas industries which suffered from chronically low rates of return tended to deteriorate their workers' earning standing.[1]

1. Figure 4 presents graphically the association of the level of wages and rates of return during the two sub-periods, 1948-51 and 1953-56. Our sample of 77 industries is divided into low-wage, medium-wage and high-wage groups (with approximately equal numbers of industries) according to average rankings of individual industries with respect to average hourly earnings in the two sub-periods. Then, each group is rearranged by their rates of return (averaged for each period). Curves in the figure above are cumulative distributions by employment for the three groups. Low-wage industries, predominantly producing consumer non-durables, had definitely lower rates of return. But medium-wage and high-wage industries were not significantly

(b) Labor productivity. To be more comparable with Census production indexes available for 1954 relative to 1947, the 1955 wage structure was compared with the 1947 wage structure. The simple correlation coefficient of the 1954/47 labor productivity rankings is -.04 with the 1947 wage structure and .03 with the 1955 wage structure. With the effect of the 1947 wage structure eliminated, the partial correlation coefficient between productivity gains and the 1955 average hourly earnings is improved to .10. But this is not significantly different from zero at the 1% level for our sample size. We may conclude that though productivity gains may have had a positive effect on variations in the relative wage structure over the period, it is not important enough to be comparable with the net association of the rate-of-return structure with variations in the interindustry wage structure.[1] Moreover, it must be noted that average rates of return were not significantly related to labor productivity changes.

(c) Concentration. Industrial concentration was expected to have a positive effect on interindustry wage variations through influencing employers' ability to capture productivity gains for their own

---

different from each other. As the high-wage group includes industries of the largest size in our sample, namely the iron and steel industry and the automobile industry, the proportions of employment among the three groups were roughly 40% (high-wage), 25% (medium-wage) and 35% (low-wage) in 1948.

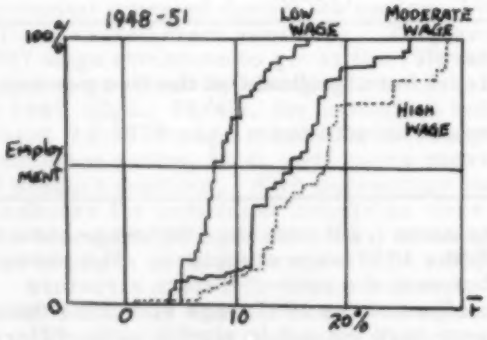


FIGURE 4A

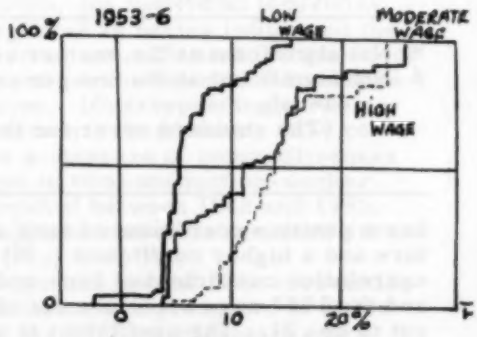


FIGURE 4B

1. This finding does not imply that changes in productivity are not an important dynamic determinant of the relative wage structure. On the contrary, we can probably demonstrate, data permiting, that differential changes in productivity are a very important factor in exercising strong influences on economic attributes, including wage rates paid, of individual firms in the same industry.

It must also be pointed out that the conventional measure of productivity --output per manhour--is not a highly relevant variable in our context. As all input prices are subject to constant change, relative changes in labor productivity alone are by no means an accurate indicator of relative cost pressures on the profitability of different industries. Productivity changes should be measured in a way different from the present practice if we really intend to measure their effects.

and their employees' benefit. Changes in output per manhour have been proven not to be a significant factor in interindustry wage variations. Yet competitiveness of product markets can be expected to remain an important variable.[1] The concentration-ratio rankings have a positive correlation both with the 1947 wage structure (.15) and the 1957 wage structure (.23). The net effect of the concentration rankings on the 1957 wage structure is measured by the partial correlation coefficient of .18, which is not significant on the 1% level but is significant at the 5% level. Highly concentrated industries tended to have somewhat higher average hourly earnings in 1947 on the average and tended to improve them faster than less concentrated industries which were originally in a similar position. To take into account this positive effect of industrial concentration on the interindustry wage structure in assessing the net effect of profitability, a combined rank order of concentration and the 1947 wage structure were correlated with the 1948-56 average rate-of-return structure. The resulting partial rank correlation coefficient, .21, is about the same as the one computed before without concentration explicitly taken into account in spite of a positive correlation (.19) between rates of return and industrial concentration.

(d) Changes in employment. Percentage changes in employment from 1948 to 1956, when ranked, turned out to be very strongly correlated with both the original and terminal wage structures and the rate-of-return structure. The rank correlation coefficients are .37, .41, and .40 respectively. The net effect of employment changes on the wage structure when the original level of earnings is taken into consideration, is measured by a much smaller coefficient of .20 (significant at the 1% level). The high correlation between changes in employment, 1948 to 1956, and the original 1947 wage structure suggests that patterns of expansion in industrial employment among manufacturing industries must have remained more or less similar before and after 1947. We suggested earlier that differential rates of growth in industrial production tend to have a substantial positive effect on relative prosperity of individual industries. The high correlation (.40) between employment changes and rates of return during the period 1948-56 reaffirms this point. A question may be raised as to whether the positive effect of interindustry wage structure is entirely due to this growth influence acting on profitability and wage changes simultaneously. To test such a hypothesis, rankings of employment changes and those of 1947 averaged hourly earnings are added together and reranked. Then, we computed a partial rank correlation coefficient between the rate-of-return structure and the 1957 wage structure with respect to this new rank order of employment changes and initial relative wages. The coefficient turned out to be .17, somewhat lower than before but yet significant at the 5% level.

4. Interpretation of Statistical Findings. Through the evaluation of these rank correlation coefficients we can observe certain broad relationships among these variables. Oligopolistic industries tend to have higher rates of return over the cycles and to pay better wages to their workers. If demand for their products is also rising at a rapid rate, they will be benefited from more favorable financial positions and will be able to afford to increase wage rates for their employees. Wage advances are necessary in any case to recruit new workers with good skill into the expanding industries. During the period, 1948-1956, changes in labor productivity seem to have had

1. Industrial concentration is an important factor in determining the long-run profitability of industries, as we saw earlier. In addition, concentration is also correlated with the extent of unionism. See Garbarino, op. cit., p. 302.

a limited effect on industries' profitability as well as on rates of wage increases. Oligopolistic firms tended to maintain their relative financial superiority with or without productivity gains. Workers in oligopolistic and usually heavy industries tend to earn high wages in any period. These industries tend to set the pace of wage advances and other more competitive and/or stagnating industries follow their patterns. If increases in wages assume the across-the-board type, low-wage competitive industries will experience above-average increases in their wage cost per unit of output.[1] Because their markets are competitive, increases in wage cost tend to deteriorate profitability of these industries. Their weakened financial standing makes employers in these industries more reluctant to grant their workers further wage increases.

If this brief discussion describes the basic mechanism which operated in American manufacturing in recent years, we should expect the following wage behavior in the post-war period. In the earlier part of the period, aggregate demand was strong and spread over all manufacturing industries more or less alike. Wages increased rapidly in all industries in more or less equal absolute amounts from year to year; the interindustry wage structure narrowed considerably. But in the later part of the period, when growth in aggregate demand slackened and influenced some industries more than others, strong oligopolistic and growing industries tended to maintain their profitability and at the same time to continue to make wage increases at similar paces as before. On the other hand, more competitive industries and those with less rapid rates of growth could not keep pace with these pattern-setting industries; both their rates of profit and wage rates tended to lag behind. Actual data are consistent with this argument.[2]

5. Conclusion. Our statistical test shows that for the period, 1947 to 1957, there was some influence of profit differentials among industries on wage differentials, while differential movements of labor productivity had little influence on differential changes in hourly earnings. Industrial concentration had a small positive effect on wage changes. The most significant factor in influencing the interindustry wage structure through this period is found to be differential changes in employment among industries. However, after all other influences are taken into account, it is to be noted that interindustry profit differentials still have some effect on changes in interindustry wage differentials.

---

1. To the extent that changes in labor productivity are associated with competitiveness of product markets, as during 1947-54, these wage cost increases are stronger.

2. A similar hypothesis was presented by Charles L. Schultze, Recent Inflation in the United States, Study Paper No. 1, Joint Economic Committee, Congress of the United States, (Washington, D. C., Government Printing Office, 1959). He sought for the explanation of the 1955-57 inflation in an uneven distribution of excess demand among industries. Should wages be responsive to excess demand and inflexible downwards, then wages continue to rise even if aggregate demand slackens as long as there remain excess demands for some industries.

TABLE 1

Statistics of Income Industry Classification, Concentration Ratios and Types of Markets

SOI = Statistics of Income Industries
SIC = Standard Industrial Classification Industries
CR = Concentration Ratios in % (weighted averages of 4-digit ratios - 4 largest companies' sales)
M = Type of Market (R-regional, L-local, otherwise national)

SOI	CR	M	SIC	Name
191	10	L	2081	Nonalcoholic beverages
192	28	R	2082, 2083	Malt and malt liquors
194	38		2084	Wines
195	64		2085	Liquors, distilled, rectified, etc.
201	27		201	Meat products
202	27		202	Dairy products
203	33		203	Canning and preserving
204	32		204 less 2043	Grain-mill products exc. cereal preparations
205	30	L	205	Bakery products
206	68		206	Sugar
207	40		207	Confectionery
208	88		2043	Cereal preparations
209	42		209	Miscellaneous food preparations
212	44		2121	Cigars
219	77		2111, 2131, 2141	Other tobacco manufactures
221	32		2211, 2212, 222	Yarn and thread mills
222	27		2213	Broadwoven fabrics (woolen & worsted)
223	18		2233	Broadwoven fabrics (cotton, etc.)
224	13		2241	Narrow fabrics
225	16		225	Knit goods
226	25		2261, 2216	Dyeing and finished textiles
227	54		227	Floor coverings
228	55		228	Hats, exc. cloth and millinery
229	33		229, 2234	Other textile-mill products
231	16		231, 232	Men's and boys' clothing
233	9		233, 234, 236	Women's clothing
235	7		2351	Millinery
237	4		2371	Fur goods
238	24		238	Other apparel and accessories
239	26		239	Other fabricated textile products
241	8	L	2411, 242	Logging camps, sawmills, planing mills, etc.
243	13		243	Millwork, etc.
244	23	R	244	Wooden containers
249	16		249	Other wood products
251	17		251, 252, 253	Household, office, public building, and professional furniture
254	18		2541	Partitions, shelvings, etc.
256	23		256	Window and door screens

SOI	CR	M	SIC	Name
259	34		259	Other furniture and fixtures
261	23		261	Pulp, paper and paperboard mills
266	28		2661, 267	Paper bags, paperboard containers and boxes
269	30		2641, 2651, 269	Pulp goods and miscellaneous converted paper products
271	18	L	2711	Newspapers
272	29		2721	Periodicals
273	25		273	Books
275	11		2751, 2761, 2771	Commercial printing incl. lithographing
279	21	L	2741, 278, 279	Other printing and publishing
281	45		281	Industrial inorganic chemicals
282	62		282	Industrial organic chemicals
283	32		283	Drugs and medicines
284	65		284	Soap and glycerin, etc.
285	29		285	Paints, varnishes, etc.
286	18		2893	Perfumes, cosmetics, etc.
287	30	R	287	Fertilizers
288	47		288	Oils and fats
289	45		286, 289 less 2893	Other chemical products
291	33		2911	Petroleum refining
299	48	R	293, 295, 299	Other petroleum and coal products
301	79		3011	Tires and inner tubes
309	33		3021, 3031, 3099	Other rubber products
311	18		3111	Leather, tanned, etc.
314	29		314	Footwear, except rubber
319	24		31 less 3111 and 314	Other leather products
321	68		3211, 322, 3231	Glass and glass products
324	31	R	3241	Cement, hydraulic
325	29	R	325	Structural clay products
326	47		326	Pottery and related products
327	36	R	327	Concrete, gypsum and plaster products
328	12		3281	Cut-stone and stone products
329	49		329	Abrasives, asbestos and other nonmetallic products
331	56		331	Blast furnaces, steelworks and rolling mills
332	29	L	332	Iron and steel foundries
333	68		333, 334, 335	Smelting and refining of nonferrous metals
336	22		3361	Nonferrous foundries
339	37		339	Other primary metal products
341	30		3411	Tin cans and other tinware
342	38		342	Cutlery, etc.
343	19	R	343	Heating apparatus and plumbing supplies
344	19		344	Fabricated structural metal products
346	15		346	Metal stamping, coating and engraving
347	20	R	3471	Lighting fixtures
348	19		348	Fabricated wire products
349	29		349	Other fabricated metal products
351	62		351	Engines and turbines

SOI	CR	M	SIC	Name
352	55		352	Agricultural machinery and tractors
353	24		353	Construction and mining machinery
354	17		354	Metalworking machinery
355	24		355	Special industry machinery
356	30		356	General industry machinery
357	64		357	Office and store machines
358	44		358	Service-industry and household machines
359	22	L	359	Machine parts and machine shops
361	50		361	Electrical generating, etc., apparatus
362	50		3621	Electrical appliances
363	48		3631	Insulated wire and cable
364	62		3641	Electrical equipment for vehicles
365	83		3651	Electric lamps
366	24		3661	Radios, radar, TV, etc.
367	78		366 less 3661	Other communication equipment
369	60		369	Other electrical machinery
372	47		372	Aircraft and parts
373	40		373	Ship and boat building and repairing
374	76		374	Railroad equipment
375	50		3751	Motor cycles, bicycles, etc.
379	38		3799	Transportation equipment, not elsewhere classified
381	74		3711,3712,3713,3715	Motor vehicles incl. bodies, truck trailers
384	74		3714,3716	Motor vehicle parts, etc.
391	--		1911,1951	Guns, howitzers, etc., inc. small arms
396	--		1961	Small arms ammunitions
399	--		192,193,194,199	Other ordnance and accessories
401	38		38 less 387	Instruments, professional, photographic, etc.
407	43		387	Watches and clocks
411	36		391	Jewelry, silverware, etc.
412	14		3961	Costume jewelry
413	8		3971	Fabricated plastic products
419	--		39 less 391,3961, 3971	Other manufacturing industries, not elsewhere classified

Not Allocables (200, 220, 230, 270, 280, 340, 350, 360)

Source: SOI Industry Classification based on files of the Internal Revenue Service Statistical Division

TABLE 2

Summary Table for Basic Industry Data

SOI Ind.	$\bar{n}$	$\bar{k}$	nia	$\bar{d}$	A_g	Q_g	Q_{sd}	$\bar{r}$	Q/L	w-47	w-57	MH 56/48	$\bar{r}$	Q/L	Ranks w_{47}	w_{57}	MH
191	6.75	68.4	4.79	14.60	5.20	--	--	9.85	140	102	163	--	71	22	65	63	--
192	9.92	69.8	.18	8.32	4.01	.17	1.33	14.58	140	--	--	--	38	--	--	--	--
194	2.71	85.9	.50	23.90	2.08	--	--	3.45	127	--	--	--	111	--	--	--	--
195	4.97	53.4	1.25	6.58	2.64	-.77	9.58	9.23	131	121	221	--	76	34	43	27 1/2	--
201	.87	16.2	.07	20.51	3.44	3.33	3.31	5.38	116	123	215	119.6	104	63 1/2	39	35	13
202	3.45	33.9	.12	4.85	3.56	1.08	3.58	10.20	147	104	184	69.0	68	17	61 1/2	56 1/2	68
203	4.16	55.2	.17	12.34	6.45	2.87	4.78	7.64	154	104	163	99.3	89	11	61 1/2	63	35
204	2.60	28.5	1.15	7.52	9.12	.56	3.12	9.07	123	108	190	--	80	48	59	49	--
205	5.11	34.7	.36	7.56	6.26	-.31	1.45	14.73	118	107	188	86.6	36	59	60	52	50
206	4.61	64.2	.25	3.97	3.26	2.15	5.56	7.14	138	109	195	88.3	93	25 1/2	58	44 1/2	48
207	6.31	41.6	.60	12.99	2.09	.35	2.20	15.42	120	103	162	76.1	29	55 1/2	63	65	60
208	9.85	42.9	.73	.76	11.52	--	--	23.05	105	--	--	--	5	--	--	--	--
209	6.30	49.5	2.84	7.68	4.80	.81	2.14	12.83	135	111	187	81.1	49	29 1/2	56 1/2	53 1/2	56
212	5.14	81.6	.43	8.67	.88	.73	2.27	6.32	139	86	132	70.9	96	24	77	77	64
219	11.19	94.4	.03	.19	3.77	.82	3.25	11.80	127	93	163	96.0	57	39	76	63	36
221	6.40	68.8	.08	11.83	4.16	--	--	9.81	145	98	138	69.3	72	18	71 1/2	76	67
222	2.58	62.6	.02	22.34	-3.49	-3.92	10.51	4.61	125	116	160	--	109	42 1/2	51	66	--
223	7.25	60.0	.00	6.82	3.97	--	--	12.74	133	101	143	--	51	31	67 1/2	73	--
224	4.16	49.6	.13	14.41	2.92	--	--	8.50	190	--	152	--	86	--	--	--	--
225	3.74	50.8	.51	13.46	7.38	1.18	5.05	7.48	130	100	145	86.9	90	35 1/2	69	72	49
226	4.48	51.4	.49	16.07	2.56	-.56	3.69	8.92	116	--	164	--	83	--	--	--	--
227	7.25	67.1	1.20	9.87	8.50	--	--	10.15	126	121	184	72.4	70	41	43	56 1/2	63
228	1.84	54.3	2.35	18.93	.27	--	--	3.41	102	--	164	--	112	--	--	--	--
229	5.50	47.6	1.12	11.19	2.18	--	--	9.51	147	--	173	--	74	--	--	--	--
231	2.59	40.4	.42	13.08	3.74	1.03	3.77	6.39	109	102	140	103.1	95	71 1/2	65	75	28
233	1.42	27.8	.11	20.17	3.38	--	--	5.11	110	126	155	104.4	107	70	34 1/2	67	24
235	-.16	22.4	.00	31.43	3.67	--	--	-.70	98*	134	173	85.4	113	76	18 1/2	60	51
237	-1.21	39.8	.00	36.84	-1.67	--	--	-3.45	114*	--	--	--	114	--	--	--	--
238	1.93	38.3	.18	18.00	5.58	--	--	5.03	119	--	139	--	108	--	--	--	--

SOI Ind.	$\bar{n}$	$\bar{k}$	nia	$\bar{d}$	A_g	Q_g	Q_{sd}	$\bar{r}$	Q/L	MH w-47	MH w-57	MH 56/48	Ranks $\bar{r}$	Ranks Q/L	Ranks w_{47}	Ranks w_{57}	Ranks MH
239	2.02	38.4	.14	16.01	6.17	--	--	5.36	127*	95	150	102.4	105	39	73 1/2	69	31
241	5.55	70.4	.22	8.93	7.81	.97	4.68	8.20	149	118	190*	83.5	87	15	49	50	54
243	5.03	50.7	.29	11.26	18.11	8.35	7.29	10.32	141	114	189	103.7	67	22	52 1/2	51	25
244	2.91	47.7	.05	12.25	-.76	-1.58	4.94	6.06	119*	94	142	73.4	98	57 1/2	74	74	62
249	3.41	54.6	.18	15.86	4.81	--	--	6.28	125*	98	152	89.2	97	42 1/2	71 1/2	68	47
251	5.78	44.4	.21	11.27	9.82	1.90	6.49	12.36	126	--	171	--	54	--	--	--	--
254	4.83	48.8	.06	12.46	10.11	--	--	11.79	136*	--	212	--	58	--	--	--	--
256	4.52	47.4	.81	15.36	8.98	--	--	9.62	191	--	--	--	73	--	--	--	--
259	4.68	44.9	.00	15.29	.95	--	--	10.51	145*	--	--	--	66	--	--	--	--
261	13.61	79.0	.15	2.20	12.94	5.29	4.66	17.72	132	122	217	108.8	12	32 1/2	40 1/2	32 1/2	22
266	8.11	47.1	.28	6.23	7.19	--	--	17.37	116	--	--	--	16	--	--	--	--
269	7.49	51.8	.15	6.56	18.86	--	--	14.53	123	--	--	--	39	--	--	--	--
271	9.19	62.8	9.22	6.31	6.08	3.44	4.11	14.64	104	175	385	112.5	37	73	1	1	17
272	4.81	47.4	8.39	13.65	3.29	--	--	10.17	186	157	252	91.6	69	1	3 1/2	5 1/2	44
273	5.10	70.8	.98	9.30	13.12	--	--	7.20	88	134	213	126.6	92	77	18 1/2	37	9
275	5.77	50.7	.44	8.92	13.37	--	--	11.40	124*	146	241	112.4	61	44 1/2	6	11	18
279	5.09	48.4	.91	11.22	6.20	--	--	10.55	123*	--	247	--	65	--	--	--	--
281	12.93	98.7	.40	1.72	5.93	8.14	3.18	12.95	138	138	244	133.5	48	25 1/2	13	9 1/2	6
282	19.93	71.7	1.22	2.10	9.49	8.33	6.84	28.69	171	131	237	120.7	2	4	26 1/2	13 1/2	12
282	16.26	75.9	.86	3.20	6.48	--	--	21.45	183	119	203	103.2	7	2	46 1/2	41	27
284	6.14	37.1	1.29	9.75	7.89	1.17	3.84	16.54	157	--	234	--	19	--	--	--	--
285	6.88	49.5	.26	3.50	7.45	2.56	5.31	13.92	121	126	218	99.9	44	53	34 1/2	31	34
286	8.00	52.5	1.68	10.94	18.46	--	--	15.36	190	--	177	--	30	--	--	--	--
287	5.87	70.6	.02	8.99	11.72	3.37	2.83	8.66	164	95	169	92.1	85	7	73 1/2	61	43
288	3.81	41.7	.43	6.97	-.87	3.37	3.75	9.22	173	99	176	66.2	77	3	70	57	69
289	10.95	69.2	.55	3.76	5.10	--	--	15.84	150	--	--	--	28	--	--	--	--
291	4.50	79.3	1.65	1.28	10.92	5.15	3.47	5.75	149	157	276	90.4	101	15	3 1/2	2	46
299	7.36	64.6	7.35	5.01	.29	.65	6.07	11.75	123	127	233	94.9	59	48	33	17	40
301	6.67	52.7	.39	1.47	10.55	2.24	7.79	13.07	124	160	263	95.6	47	44 1/2	2	4	38
309	7.77	51.2	.60	6.43	12.41	4.56	7.43	14.88	137	120	194	111.9	33	27	45	46	19
311	2.53	47.6	.37	19.96	-3.05	-.15	6.25	5.34	109	124	195	77.8	106	71 1/2	37 1/2	44 1/2	57
314	3.71	39.8	.13	9.74	4.48	1.20	4.29	9.29	102	101	149	95.9	75	75	65	70	37
319	2.56	37.7	.47	18.21	5.89	-.69	4.38	6.77	112	101	148	95.5	94	69	67 1/2	71	39
321	12.70	58.5	.04	3.99	14.07	4.83	6.17	21.95	126	--	224	--	6	--	--	--	--
324	24.33	102.0	.01	.40	21.57	5.64	2.01	23.90	127	118	216	103.3	4	39	49	34	26
325	10.54	71.6	.15	5.06	8.11	3.13	6.05	15.03	117	111	187	102.7	32	61	56 1/2	53 1/2	29

SOI Ind.	$\bar{n}$	$\bar{k}$	nia	$\bar{d}$	A_g	Q_g	Q_{sd}	$\bar{r}$	Q/L	w-47	w-57	MH 56/48	$\bar{r}$	Q/L	w_{47}	w_{57}	MH
326	6.17	57.1	.30	12.34	4.58	1.09	4.98	10.98	116	118	197	83.7	63	63 1/2	49	43	53
327	11.69	67.8	.09	4.40	20.98	8.54	7.62	17.49	157	114	192	119.0	13	9	52 1/2	47	15
328	3.69	63.5	.29	11.38	2.22	--	--	5.89	109*	--	177	--	100	--	--	--	--
329	10.50	64.2	.07	2.32	8.68	--	--	16.47	126	--	214	--	20	--	--	--	--
331	10.32	66.6	.02	1.27	9.18	3.94	9.78	15.16	115	144	268	101.7	31	66	8 1/2	3	32
332	8.52	53.7	.31	7.31	7.12	1.49	9.81	16.10	115	133	223	93.4	26	66	21 1/2	24 1/2	41
333	8.75	73.4	.16	3.41	16.47	4.43	7.63	12.39	128	130	237	119.6	53	37	30 1/2	13 1/2	14
336	5.08	43.4	.36	13.99	6.45	4.26	11.12	11.82	103	137	228	83.4	56	74	15 1/2	19	55
339	9.89	54.0	.10	4.53	11.29	--	--	18.54	105	--	249	--	10	--	--	--	--
341	6.63	63.7	.09	.45	13.36	4.86	3.80	10.67	117	119	234	114.8	64	61	46 1/2	16	16
342	10.43	61.0	1.21	7.26	5.12	--	--	17.24	122	121	212	91.5	17	51	43	38 1/2	45
343	7.63	55.4	.10	6.20	4.29	-.12	13.19	13.86	140	131	212	75.9	45	22	26 1/2	38 1/2	61
344	7.49	50.8	.18	6.42	14.49	--	--	14.73	123	130	223	134.7	35	48	30 1/2	24 1/2	4
346	7.07	49.7	.27	9.03	10.93	--	--	14.32	113*	129	222	132.7	41	68	32	26	7
347	3.70	40.4	.15	12.94	11.15	--	--	9.22	110*	--	201	--	78	--	--	--	--
348	8.58	51.0	.08	5.45	7.79	--	--	17.03	113	--	205	--	18	--	--	--	--
349	8.85	54.9	.15	8.31	13.34	--	--	16.38	104	--	215	--	21	--	--	--	--
351	8.72	60.1	.23	4.60	9.06	--	--	14.51	141	144	244	105.4	40	20	8 1/2	9 1/2	23
352	8.48	66.8	.07	3.57	7.38	-1.85	9.57	12.76	130	137	230	69.9	50	35 1/2	15 1/2	18	66
353	11.88	65.8	.16	4.23	13.87	--	--	18.17	119	131	227	123.9	11	57 1/2	26 1/2	20	10
354	10.28	70.6	.36	7.28	15.27	8.49	17.36	14.85	117	139	249	134.6	34	61	11	7	5
355	8.44	73.0	.14	6.56	7.01	--	--	12.60	121	131	217	85.0	52	53	26 1/2	32 1/2	52
356	9.30	59.0	.62	4.64	16.18	--	--	16.22	123	134	226	110.8	25	48	18 1/2	21	21
357	13.34	82.0	.20	2.30	14.84	--	--	16.30	155	138	225	102.6	22	10	13	22	30
358	7.52	56.7	.46	10.13	12.57	4.71	14.37	13.57	151	134	221	101.3	46	13	18 1/2	27 1/2	33
359	9.80	61.0	.16	8.62	14.97	--	--	16.23	123	132	224	149.9	24	48	23	23	2
361	11.06	53.4	.12	2.79	20.25	--	--	20.69	136	133	219	122.2	8	28	21 1/2	29 1/2	11
362	8.87	51.5	.05	8.08	14.45	--	--	17.39	159	--	212	--	15	--	--	--	--
363	7.84	46.2	.00	2.60	14.92	--	--	17.41	140*	--	205	--	14	--	--	--	--
364	8.63	54.7	.00	6.43	15.01	--	--	15.92	135	131	219	111.0	27	29 1/2	26 1/2	29 1/2	20
365	4.41	50.1	.00	9.40	11.97≠	--	--	8.86	141	--	193	--	84	--	--	--	--
366	6.43	46.5	.12	8.10	37.48	11.40	25.85	14.27	170	113	191	--	42	5	54	48	--
367	7.22	63.7	1.40	4.58	-3.82	--	--	11.46	167	125	207	--	60	6	36	40	--
369	4.72	52.3	3.89	6.64	4.51	--	--	9.19	120	--	202	--	79	--	--	--	--
372	6.27	59.6	.13	5.32	56.20	26.76	22.24	11.15	159*	138	236	318.3	62	8	13	15	1
373	4.39	71.7	.35	16.19	5.77	2.72	15.84	6.02	142	145	239	92.8	99	19	7	12	42

SOI Ind.	$\bar{n}$	$\bar{k}$	nia	$\bar{d}$	A_g	Q_g	Q_{sd}	$\bar{r}$	Q/L	w-47	w-57	MH 56/48	$\bar{r}$	Q/L	Ranks w_{47}	w_{57}	MH
374	6.00	76.0	.55	1.70	3.60	-8.73	19.90	8.02	120	141	252	70.6	88	55 1/2	10	5 1/2	65
375**	2.46	49.0	.62	10.43	2.36	--	--	5.47	137	131	202	50.8	102	32 1/2	26 1/2	42	70
379	3.01	56.1	.00	15.14	.66	--	--	5.45	116*				103				
381**	12.56	42.0	.64	4.82	15.07	4.33	13.45	31.13	121	147	246	136.0	1	53	5	8	3
384	9.38	49.4	.11	5.44	7.00	2.40	10.28	19.17					9				
391	9.82	68.1	.00	5.58	-1.09	--	--	14.29	--	--	--	--	43	--	--	--	--
396	17.37	74.3	.00	.65	-7.64	--	--	25.27	--	--	--	--	3	--	--	--	--
399	3.54	57.3	.00	16.76	36.49	--	--	7.46	--	--	--	--	91	--	--	--	--
401	10.51	64.6	.50	4.48	24.17	--	--	16.27	149	124	214	126.8	23	15	37 1/2	36	8
407	6.67	65.7	.05	10.17	1.08	--	--	12.06	152	112	185	76.4	55	12	55	55	58
411	5.02	56.1	.13	12.80	3.65	--	--	9.03	115	122	182	76.2	81	66	40 1/2	58	59
412	1.48	37.3	.27	24.62	26.26	--	--	3.90	120*	--	--	--	110	--	--	--	--
413	4.04	45.5	.36	14.82	21.45	--	--	8.94	171*	--	191	--	82	--	--	--	--

Notations:

$\bar{n}$	Percentage net margins, averaged over the period, 1948-1956, exc. 1952 (in per cent)
$\bar{k}$	Assets/Sales ratios, averaged over the period (in per cent)
$\bar{d}$	Net-deficit sales ratios, averaged over the period (in per cent)
nia	Net intangible assets as percentage of total capital invested in 1956
A_g	Arithmetic rates of change in capital invested between 1948 and 1956 (in per cent); ≠ between 1951 and 1956 for Industry 281, between 1950 and 1956 for Industry 365
Q_g	Trend rates of growth in output between 1947 and 1957 (in per cent)
Q_{sd}	Standard deviations of the actual/trend output ratios (in per cent)
$\bar{r}$	Rates of return (before tax) on capital invested, averaged over the period (in per cent)
Q/L	Output per production-worker manhour, 1954 relative to 1947; * "indirectly represented" output indexes
w-47	Average hourly earnings (in cent), production worker manhour, 1947
w-57	Average hourly earnings (in cent), production worker manhour, 1957; *w for Industry 241 is extrapolated from 1956 data
MH, 56/48	Total production-worker manhours, 1956 relative to 1947 (in per cent)
Ranks $\bar{r}$	Rankings of $\bar{r}$, 114 industries
Q/L	Rankings of Q/L, 77 industries
w-47	Rankings of w-47, 77 industries
w-57	Rankings of w-57, 77 industries
MH	Rankings of MH, 69 industries
**	$\bar{r}$ and Q/L for Industry 375 and 379 combined are 5.46 and 132. $\bar{r}$ for 381 and 384 combined is 29.14.